OM protein - protein search, using sw model

Run on: May 4, 1999, 12:32:39 ; Search time 27.23 Seconds
(without alignments)
11.141 Million cell updates/sec

Title: US-09-037-460-2_COPY_55_69
Perfect score: 78
Sequence: 1 RVCAAGRGETCYRTV 15

Scoring table: PAM150

Searched: 162890 seqs, 20225328 residues

Database : A_Geneseq_34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	78	100.0	184	1	R98994	Vascular IBP-like
2	34	43.6	18	1	R57843	VnRbeta3-7, bindin
3	33	42.3	18	1	R57842	VnRbeta3-4, bindin
4	35	44.9	45	1	W10165	Alpha-hordothionin
5	35	44.9	50	1	W76687	Elapidae modified
6	36	46.2	75	1	R31601	Chicken nov protei
7	36	46.2	76	1	R31600	Chicken nov protei
8	35	44.9	60	1	W76646	Elapidae modified
9	35	44.9	67	1	W76686	Elapidae modified
10	35	44.9	70	1	W76685	Elapidae modified
11	35	44.9	72	1	W76684	Elapidae modified
12	35	44.9	73	1	W76658	Elapidae modified
13	35	44.9	74	1	W76661	Elapidae modified
14	30	38.5	11	1	R89364	Elapidae modified
15	31	39.7	24	1	W03034	MAGE-1 derived imm
16	31	39.7	25	1	R32778	Thrombolytic enzym
17	31	39.7	25	1	R37753	MVIIB omega conoto
18	31	39.7	25	1	R39609	MVIIB/SNX-159. Red
19	31	39.7	25	1	R76090	MVIIB/SNX159. Use
20	31	39.7	25	1	W12968	Omega conotoxin MV
21	31	39.7	25	1	W19545	Omega conopeptide
22	31	39.7	25	1	W72606	Natural omega-cono
23	30	38.5	18	1	R57849	Conus genus natura
24	31	39.7	28	1	W07695	MTFVbeta3-18, bind
25	29	37.2	14	1	R47977	Dendroides canaden
26	29.5	37.8	17	1	P91671	Lactobacillus brev
27	29.5	37.8	17	1	R06266	New lipopolysaccha
28	29.5	37.8	17	1	R06861	Antiviral peptide.
29	29.5	37.8	17	1	R06862	Tachyplesin I. Wat
30	29.5	37.8	17	1	R08202	Tachyplesin II. Wa
31	29.5	37.8	17	1	R23112	Gigasin II. Novel
32	29.5	37.8	17	1	R23113	Bacterial shock tr
33	29.5	37.8	17	1	R23114	Bacterial shock tr
34	29.5	37.8	17	1	R38491	Bacterial shock tr
35	29.5	37.8	17	1	R38490	Tacyplesin-III. Be
36	29.5	37.8	17	1	R38489	Tacyplesin-II. Bet
37	29.5	37.8	17	1	R75805	Tacyplesin-I. Beta
38	29.5	37.8	17	1	W66465	Tachyplesin, an an
39	29.5	37.8	17	1	W66466	Cationic peptide t
40	28	35.9	10	1	W52077	Cationic peptide t
41	38	48.7	504	1	R32710	Targetting peptide
42	28	35.9	11	1	P91265	Haematopoietic ste
43	29	37.2	17	1	W78913	Tissue plasminogen
44	28	35.9	12	1	W52102	Rat cocaine and am
45	29	37.2	18	1	R57840	Targetting peptide VnR-beta3-5 (Fab9)

ALIGNMENTS

```
RESULT   1
R98994
ID   R98994 standard; Protein; 184 AA.
AC   R98994;
DT   06-NOV-1996  (first entry)
DE   Vascular IBP-like growth factor.
KW   Vascular IBF-like growth factor; VIGF;
KW   insulin-like growth factor binding protein; agonist; antagonist;
KW   muscle wastage; osteoporosis; implant fixation; wound healing;
KW   therapy; diagnosis.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1. .21
FT                   /label= Sig_peptide
PN   WO9617931-A1.
PD   13-JUN-1996.
PF   09-DEC-1994; U14388.
PR   09-DEC-1994; WO-U14388.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Hastings GA,  Rosen CA;
DR   WPI; 96-287176/29.
DR   N-PSDB; T34991.
PT   Human vascular insulin-like growth factor binding protein-like
PT   growth factor, and its nucleic acid sequence and (ant)agonists -
PT   used, e.g. to treat muscle wasting diseases or aid implant fixation,
PT   or limit excess connective tissue prodn. during wound healing.
PS   Claim 14; Page 43-44; 61pp; English.
CC   Human vascular insulin-like growth factor binding protein-like
CC   growth factor (R98994), or VIGF, is a protein of primarily
CC   vascular origin that is structurally related to the IBP and CCN
CC   protein families.  It can be expressed in e.g. E. coli, CHO or
CC   insect host cells using a vector incorporating a cDNA clone
CC   (T34991), or its derivative, obtd. from human umbilical
CC   endothelial cells.  It is useful therapeutically e.g. for
CC   treating muscle wasting diseases or osteoporosis, or can be used
CC   to detect diseases associated with under- or over-expression of VIGF,
CC   or to screen for antagonists useful during wound healing.
SQ   Sequence   184 AA;

  Query Match           100.0%;  Score 78;  DB 1;  Length 184;
  Best Local Similarity  100.0%;  Pred. No. 7e-05;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 RVCAAGRGETCYRTV 15
           |||||||||||||||
Db      55 RVCAAGRGETCYRTV 69


RESULT   2
R57843
ID   R57843 standard; peptide; 18 AA.
AC   R57843;
DT   28-MAR-1995  (first entry)
DE   VnRbeta3-7, binding site for vitronectin receptor alpha-v, beta-3.
KW   Binding site; CDR; complementarity determining region; immunoglobulin;
KW   heavy; light; primer extension; PCR; amplify; fibronectin; vitronectin;
KW   RGD-dependant; integrin ligand; von Willebrand factor; EBV; gp350/220;
KW   envelope gylcoprotein; HIV; gp120; reovirus; hemagglutinin; insulin;
KW   cellular receptor; CR2; CD4; hormone; thyroid stimulating hormone; TSH;
KW   transferrin; apolipoprotein; apo E; apo AI; MHC; class I; class II;
KW   non-RGD-dependent; vitronectin receptor; alpha-v, beta-3; modulation;
KW   anti-gpIIb/IIIa; monoclonal antibody; MAb; platelet adhesion; cancer;
KW   coagulation; inflammation; anti-vitronectin; tumour cell adhesion;
KW   migration.
OS   Homo sapiens.
```

PN WO9418221-A.
PD 18-AUG-1994.
PF 02-FEB-1994; U01258.
PR 02-FEB-1993; US-012566.
PR 28-JUN-1993; US-084542.
PA (SCRI) SCRIPPS RES INST.
PI Barbas CF, Lerner RA;
DR WPI; 94-279675/34.
PT Production of binding sites within CDR regions of immunoglobulins
PT - displayed on the surface of filamentous phage particles, for
PT inhibiting platelet aggregation and vitronectin binding
PS Claim 44; Page 22; 207pp; English.
CC The sequences given in R57837-84 are binding sites which were used in
CC the method of the invention for producing a polypeptide having a
CC binding site capable of binding a preselected agent. Nucleotide
CC sequences encoding these binding site peptides were introduced into
CC a CDR region of a nucleic acid encoding an immunoglobulin heavy (H)
CC or light (L) chain, by amplifying the CDR region by primer extension.
CC Preferred binding sites are derived from the RGD-dependant integrin
CC ligands, eg. fibronectin, vitronectin, von Willebrand factor, from
CC the envelope gylcoprotein from viruses such as HIV gp120, EBV gp350/
CC 220, reovirus hemagglutinin, from cellular receptors such as CR2 or
CC CD4, from protein hormones such as thyroid stimulating hormone (TSH),
CC insulin, transferrin, from apolipoproteins such as apo E and apo AI,
CC from immunoglobulin CDRs and from MHC class I or II proteins. Non-RGD-
CC dependent integrin binding sites were selected for the affinity to bind
CC vitronectin receptor alpha-v, beta-3. An anti-gpIIb/IIIa monoclonal
CC antibody (MAb) produced in this way can be used to modulate platelet
CC adhesion in the treatment of coagulation and some inflammatory responses.
CC An anti-vitronectin MAb can be used in the treatment of cancer by
CC blocking tumour cell adhesion and migration. This sequence represents
CC an RGD-dependant binding site which has been shown to bind the human
CC vitronectin receptor (VnR) alpha-v, beta-3 when present in a phagemid
CC display protein.
SQ Sequence 18 AA;

Query Match 43.6%; Score 34; DB 1; Length 18;
Best Local Similarity 58.3%; Pred. No. 27;
Matches 7; Conservative 2; Mismatches 1; Indels 2; Gaps 1;

Qy 3 CAAGRGET--CY 12
|: |||:| ||
Db 3 CTYGRGDTRNCY 14

RESULT 3
R57842
ID R57842 standard; peptide; 18 AA.
AC R57842;
DT 28-MAR-1995 (first entry)
DE VnRbeta3-4, binding site for vitronectin receptor alpha-v, beta-3.
KW Binding site; CDR; complementarity determining region; immunoglobulin;
KW heavy; light; primer extension; PCR; amplify; fibronectin; vitronectin;
KW RGD-dependant; integrin ligand; von Willebrand factor; EBV; gp350/220;
KW envelope gylcoprotein; HIV; gp120; reovirus; hemagglutinin; insulin;
KW cellular receptor; CR2; CD4; hormone; thyroid stimulating hormone; TSH;
KW transferrin; apolipoprotein; apo E; apo AI; MHC; class I; class II;
KW non-RGD-dependent; vitronectin receptor; alpha-v, beta-3; modulation;
KW anti-gpIIb/IIIa; monoclonal antibody; MAb; platelet adhesion; cancer;
KW coagulation; inflammation; anti-vitronectin; tumour cell adhesion;
KW migration.
OS Homo sapiens.
PN WO9418221-A.
PD 18-AUG-1994.
PF 02-FEB-1994; U01258.
PR 02-FEB-1993; US-012566.
PR 28-JUN-1993; US-084542.
PA (SCRI) SCRIPPS RES INST.
PI Barbas CF, Lerner RA;
DR WPI; 94-279675/34.
PT Production of binding sites within CDR regions of immunoglobulins
PT - displayed on the surface of filamentous phage particles, for
PT inhibiting platelet aggregation and vitronectin binding
PS Claim 44; Page 22; 207pp; English.
CC The sequences given in R57837-84 are binding sites which were used in
CC the method of the invention for producing a polypeptide having a
CC binding site capable of binding a preselected agent. Nucleotide
CC sequences encoding these binding site peptides were introduced into
CC a CDR region of a nucleic acid encoding an immunoglobulin heavy (H)
CC or light (L) chain, by amplifying the CDR region by primer extension.
CC Preferred binding sites are derived from the RGD-dependant integrin
CC ligands, eg. fibronectin, vitronectin, von Willebrand factor, from
CC the envelope gylcoprotein from viruses such as HIV gp120, EBV gp350/
CC 220, reovirus hemagglutinin, from cellular receptors such as CR2 or
CC CD4, from protein hormones such as thyroid stimulating hormone (TSH),
CC insulin, transferrin, from apolipoproteins such as apo E and apo AI,
CC from immunoglobulin CDRs and from MHC class I or II proteins. Non-RGD-
CC dependent integrin binding sites were selected for the affinity to bind
CC vitronectin receptor alpha-v, beta-3. An anti-gpIIb/IIIa monoclonal
CC antibody (MAb) produced in this way can be used to modulate platelet
CC adhesion in the treatment of coagulation and some inflammatory responses.
CC An anti-vitronectin MAb can be used in the treatment of cancer by
CC blocking tumour cell adhesion and migration. This sequence represents
CC an RGD-dependant binding site which has been shown to bind the human
CC vitronectin receptor (VnR) alpha-v, beta-3 when present in a phagemid
CC display protein.
SQ Sequence 18 AA;

Query Match 42.3%; Score 33; DB 1; Length 18;
Best Local Similarity 50.0%; Pred. No. 38;
Matches 6; Conservative 3; Mismatches 1; Indels 2; Gaps 1;

Qy 3 CAAGRGE--TCY 12
|: |||: :||
Db 3 CTQGRGDWRSCY 14

RESULT 4
W10165
ID W10165 standard; protein; 45 AA.
AC W10165;
DT 15-JUL-1997 (first entry)
DE Alpha-hordothionin threonine rich amino acid sequence.
KW Animal feed; food; barley.
OS Hordeum vulgare.
OS Synthetic.
PN WO9638562-A1.
PD 05-DEC-1996.
PF 31-MAY-1996; U08219.
PR 02-JUN-1995; US-459180.
PA (PION-) PIONEER HI-BRED INT INC.
PI Rao GA;
DR WPI; 97-034375/03.
PT New modified alpha-hordothionin having threonine amino acid substns.
PT - to increase the threonine content of e.g. animal feed
PS Claim 1; Page 11; 19pp; English.
CC The present sequence is a threonine rich alpha-hordothionin amino
CC acid sequence. The protein contains a threonine residue at positions
CC 1, 5, 7, 8, 11, 15, 17, 19, 22, 23, 24, 30, 32, 34, 38, and 41. The
CC threonine has substituted polar, charged and hydrophobic residues.
CC With exception of arginine at position 10, serine at position 2 and
CC lysine at position 45, as they are required for maintaining the structure
CC of the protein through a hydrogen-bonding network. The protein produced
CC can be used in foods or feeds to provide higher levels of essential amino
CC acid threonine.
SQ Sequence 45 AA;

Query Match 44.9%; Score 35; DB 1; Length 45;
Best Local Similarity 50.0%; Pred. No. 43;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 3 CAAGRGETCYRT 14
|:: | ||| |

```
Db      4 CTSTTGKTCYNT 15


RESULT   5
W76687
ID   W76687 standard; Protein; 50 AA.
AC   W76687;
DT   05-JAN-1999  (first entry)
DE   Elapidae modified dendroaspin protein fragment Den-Tb.
KW   Dendroaspin; snake venom; clotting cascade; anticoagulant; platelet;
KW   integrin binding; injury; blood; cell migration; thrombosis; inhibitor;
KW   proliferation; signal transduction; regulator; coagulation; treatment;
KW   prophylactic; artery; vein; wall thickening; myocardial infarction;
KW   retinal neovascularisation; dysregulated apoptosis; tumorigenesis;
KW   leukocyte recruitment, immune system; tissue fibrosis.
OS   Elapidae.
OS   Synthetic.
PN   WO9842834-A1.
PD   01-OCT-1998.
PF   20-MAR-1998; G00848.
PR   20-MAR-1997; GB-005787.
PA   (THRO-) THROMBOSIS RES INST.
PI   Authi K, Kakkar V, Lu X, Scully MF;
DR   WPI; 98-542278/46.
PT   New hybrid dendroaspin polypeptide(s) - used for treating, e.g.
PT   thrombosis, myocardial infarction, dysregulated apoptosis, abnormal
PT   cell migration and immune system activation
PS   Claim 6; Fig 3C; 59pp; English.
CC   W76645-W76688 represent modified dendroaspin protein fragments isolated
CC   from snake venom. When dendroaspin is modified to incorporate further
CC   functional amino acid sequence, e.g. active portions or motifs of
CC   agonists, antagonists or inhibitors of factors in the clotting cascade,
CC   the resulting molecules are particularly useful as anticoagulants. The
CC   molecules have an integrin binding activity which when administered in
CC   vivo results in the binding of the molecules to platelets thereby
CC   inhibiting the aggregation of the platelets at sites of injury. Non-wild
CC   type dendroaspin domains provide secondary, optionally further
CC   functionality, e.g. antithrombotic action, inhibiting cell migration and
CC   proliferation and regulating signal transduction. Such variants have bi-
CC   or multifunctional activities against blood coagulation, particularly
CC   thrombus formation and arterial/venous wall thickening at the sites of
CC   injury. The variants may have activities against leukocyte recruitment,
CC   immune system activation, tissue fibrosis and tumourigenesis. The
CC   polypeptides can be used for the treatment or prophylaxis of a disease
CC   associated with thrombosis, e.g. myocardial infarction, retinal
CC   neovascularisation, endothelial injury, dysregulated apoptosis, abnormal
CC   cell migration, leukocyte recruitment, immune system activation, tissue
CC   fibrosis or tumorigenesis.
SQ   Sequence   50 AA;


  Query Match             44.9%;  Score 35;  DB 1;  Length 50;
  Best Local Similarity   22.2%;  Pred. No. 47;
  Matches    6;  Conservative    6;  Mismatches    3;  Indels   12;  Gaps    1;

Qy       1 RVCAAGRG----------ETCYRTV 15
           |:| : ||          ::||: :
Db       1 RICFTPRGDMPGPYPGPCQEDSCYKNI 27


RESULT   6
R31601
ID   R31601 standard; Protein; 75 AA.
AC   R31601;
DT   24-MAY-1993  (first entry)
DE   Chicken nov protein fragment V.
KW   avian nephroblastoma; avian myeloblastoma virus;
KW   stringent hybridisation.
OS   Gallus domesticus.
PN   WO9300430-A.
PD   07-JAN-1993.
PF   25-JUN-1992; F00589.
PR   25-JUN-1991; FR-007807.
PA   (CNRS ) CENT NAT RECH SCI.
PI   Martinerie C,  Perbal B;
DR   WPI; 93-036377/04.
PT   Nucleotide sequences hybridising to regions of chicken nov gene -
PT   useful as probes for detecting complementary sequences to
PT   evaluate development and/or differentiation of tumours
PS   Claim 5; Page 28; 67pp; French.
CC   The chicken nov gene is stimulated in avian nephroblastoma induced
CC   by avian myeloblastoma virus but not in normal adult kidney. A
CC   1975bp cDNA sequence (Q36031) was isolated from a gene bank prepared
CC   from chicken embryonic fibroblasts screened with a tumour-derived
CC   probe. Nucleotide sequences which hybridise to Q36031 or specified
CC   sub-fragments of it, under stringent conditions (i.e. 50% formamide,
CC   5 x SCC), are claimed. The claimed sequences preferably encode a
CC   protein with amino acid sequence V (R31601).
SQ   Sequence   75 AA;


  Query Match             46.2%;  Score 36;  DB 1;  Length 75;
  Best Local Similarity   70.0%;  Pred. No. 48;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 VCAAGRGETC 11
           |||  |||:|
Db      38 VCARQRGESC 47


RESULT   7
R31600
ID   R31600 standard; Protein; 76 AA.
AC   R31600;
DT   24-MAY-1993  (first entry)
DE   Chicken nov protein fragment IV.
KW   avian nephroblastoma; avian myeloblastoma virus;
KW   stringent hybridisation.
OS   Gallus domesticus.
PN   WO9300430-A.
PD   07-JAN-1993.
PF   25-JUN-1992; F00589.
PR   25-JUN-1991; FR-007807.
PA   (CNRS ) CENT NAT RECH SCI.
PI   Martinerie C,  Perbal B;
DR   WPI; 93-036377/04.
DR   N-PSDB; Q36033.
PT   Nucleotide sequences hybridising to regions of chicken nov gene -
PT   useful as probes for detecting complementary sequences to
PT   evaluate development and/or differentiation of tumours
PS   Claim 3; Page 27; 67pp; French.
CC   The chicken nov gene is stimulated in avian nephroblastoma induced
CC   by avian myeloblastoma virus but not in normal adult kidney. A
CC   1975bp cDNA sequence was isolated from a gene bank prepared from
CC   chicken embryonic fibroblasts screened with a tumour-derived probe.
CC   Fragment III (Q36033) is derived from the 2nd. exon of the nov gene;
CC   nucleotide sequences which hybridise to Fragment III under stringent
CC   conditions (i.e. 50% formamide, 5 x SCC) are claimed. The claimed
CC   sequences preferably encode a protein with at least 70% homology to
CC   amino acid sequence IV (R31600) which is encoded by Fragment III.
SQ   Sequence   76 AA;


  Query Match             46.2%;  Score 36;  DB 1;  Length 76;
  Best Local Similarity   70.0%;  Pred. No. 48;
  Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 VCAAGRGETC 11
           |||  |||:|
Db      39 VCARQRGESC 48


RESULT   8
W76646
ID   W76646 standard; Protein; 60 AA.
AC   W76646;
```

DT 05-JAN-1999 (first entry)
DE Elapidae modified dendroaspin protein fragment DEN-RGD1.
KW Dendroaspin; snake venom; clotting cascade; anticoagulant; platelet;
KW integrin binding; injury; blood; cell migration; thrombosis; inhibitor;
KW proliferation; signal transduction; regulator; coagulation; treatment;
KW prophylactic; artery; vein; wall thickening; myocardial infarction;
KW retinal neovascularisation; dysregulated apoptosis; tumorigenesis;
KW leukocyte recruitment, immune system; tissue fibrosis.
OS Elapidae.
OS Synthetic.
PN WO9842834-A1.
PD 01-OCT-1998.
PF 20-MAR-1998; G00848.
PR 20-MAR-1997; GB-005787.
PA (THRO-) THROMBOSIS RES INST.
PI Authi K, Kakkar V, Lu X, Scully MF;
DR WPI; 98-542278/46.
PT New hybrid dendroaspin polypeptide(s) - used for treating, e.g.
PT thrombosis, myocardial infarction, dysregulated apoptosis, abnormal
PT cell migration and immune system activation
PS Claim 6; Fig 3A; 59pp; English.
CC W76645-W76688 represent modified dendroaspin protein fragments isolated
CC from snake venom. When dendroaspin is modified to incorporate further
CC functional amino acid sequence, e.g. active portions or motifs of
CC agonists, antagonists or inhibitors of factors in the clotting cascade,
CC the resulting molecules are particularly useful as anticoagulants. The
CC molecules have an integrin binding activity which when administered in
CC vivo results in the binding of the molecules to platelets thereby
CC inhibiting the aggregation of the platelets at sites of injury. Non-wild
CC type dendroaspin domains provide secondary, optionally further
CC functionality, e.g. antithrombotic action, inhibiting cell migration and
CC proliferation and regulating signal transduction. Such variants have bi-
CC or multifunctional activities against blood coagulation, particularly
CC thrombus formation and arterial/venous wall thickening at the sites of
CC injury. The variants may have activities against leukocyte recruitment,
CC immune system activation, tissue fibrosis and tumourigenesis. The
CC polypeptides can be used for the treatment or prophylaxis of a disease
CC associated with thrombosis, e.g. myocardial infarction, retinal
CC neovascularisation, endothelial injury, dysregulated apoptosis, abnormal
CC cell migration, leukocyte recruitment, immune system activation, tissue
CC fibrosis or tumorigenesis.
SQ Sequence 60 AA;

Query Match 44.9%; Score 35; DB 1; Length 60;
Best Local Similarity 22.2%; Pred. No. 55;
Matches 6; Conservative 6; Mismatches 3; Indels 12; Gaps 1;

Qy 1 RVCAAGRG-----ETCYRTV 15
|:| : || ::||: :
Db 1 RICFTPRGDMPGPYPGPCQEDSCYKNI 27

RESULT 9
W76686
ID W76686 standard; Protein; 67 AA.
AC W76686;
DT 05-JAN-1999 (first entry)
DE Elapidae modified dendroaspin protein fragment Den-Gb.
KW Dendroaspin; snake venom; clotting cascade; anticoagulant; platelet;
KW integrin binding; injury; blood; cell migration; thrombosis; inhibitor;
KW proliferation; signal transduction; regulator; coagulation; treatment;
KW prophylactic; artery; vein; wall thickening; myocardial infarction;
KW retinal neovascularisation; dysregulated apoptosis; tumorigenesis;
KW leukocyte recruitment, immune system; tissue fibrosis.
OS Elapidae.
OS Synthetic.
PN WO9842834-A1.
PD 01-OCT-1998.
PF 20-MAR-1998; G00848.
PR 20-MAR-1997; GB-005787.
PA (THRO-) THROMBOSIS RES INST.
PI Authi K, Kakkar V, Lu X, Scully MF;
DR WPI; 98-542278/46.
PT New hybrid dendroaspin polypeptide(s) - used for treating, e.g.
PT thrombosis, myocardial infarction, dysregulated apoptosis, abnormal
PT cell migration and immune system activation
PS Claim 6; Fig 3C; 59pp; English.
CC W76645-W76688 represent modified dendroaspin protein fragments isolated
CC from snake venom. When dendroaspin is modified to incorporate further
CC functional amino acid sequence, e.g. active portions or motifs of
CC agonists, antagonists or inhibitors of factors in the clotting cascade,
CC the resulting molecules are particularly useful as anticoagulants. The
CC molecules have an integrin binding activity which when administered in
CC vivo results in the binding of the molecules to platelets thereby
CC inhibiting the aggregation of the platelets at sites of injury. Non-wild
CC type dendroaspin domains provide secondary, optionally further
CC functionality, e.g. antithrombotic action, inhibiting cell migration and
CC proliferation and regulating signal transduction. Such variants have bi-
CC or multifunctional activities against blood coagulation, particularly
CC thrombus formation and arterial/venous wall thickening at the sites of
CC injury. The variants may have activities against leukocyte recruitment,
CC immune system activation, tissue fibrosis and tumourigenesis. The
CC polypeptides can be used for the treatment or prophylaxis of a disease
CC associated with thrombosis, e.g. myocardial infarction, retinal
CC neovascularisation, endothelial injury, dysregulated apoptosis, abnormal
CC cell migration, leukocyte recruitment, immune system activation, tissue
CC fibrosis or tumorigenesis.
SQ Sequence 67 AA;

Query Match 44.9%; Score 35; DB 1; Length 67;
Best Local Similarity 22.2%; Pred. No. 61;
Matches 6; Conservative 6; Mismatches 3; Indels 12; Gaps 1;

Qy 1 RVCAAGRG-----ETCYRTV 15
|:| : || ::||: :
Db 1 RICFTPRGDMPGPYPGPCQEDSCYKNI 27

RESULT 10
W76685
ID W76685 standard; Protein; 70 AA.
AC W76685;
DT 05-JAN-1999 (first entry)
DE Elapidae modified dendroaspin protein fragment Den-Tm.
KW Dendroaspin; snake venom; clotting cascade; anticoagulant; platelet;
KW integrin binding; injury; blood; cell migration; thrombosis; inhibitor;
KW proliferation; signal transduction; regulator; coagulation; treatment;
KW prophylactic; artery; vein; wall thickening; myocardial infarction;
KW retinal neovascularisation; dysregulated apoptosis; tumorigenesis;
KW leukocyte recruitment, immune system; tissue fibrosis.
OS Elapidae.
OS Synthetic.
PN WO9842834-A1.
PD 01-OCT-1998.
PF 20-MAR-1998; G00848.
PR 20-MAR-1997; GB-005787.
PA (THRO-) THROMBOSIS RES INST.
PI Authi K, Kakkar V, Lu X, Scully MF;
DR WPI; 98-542278/46.
PT New hybrid dendroaspin polypeptide(s) - used for treating, e.g.
PT thrombosis, myocardial infarction, dysregulated apoptosis, abnormal
PT cell migration and immune system activation
PS Claim 6; Fig 3C; 59pp; English.
CC W76645-W76688 represent modified dendroaspin protein fragments isolated
CC from snake venom. When dendroaspin is modified to incorporate further
CC functional amino acid sequence, e.g. active portions or motifs of
CC agonists, antagonists or inhibitors of factors in the clotting cascade,
CC the resulting molecules are particularly useful as anticoagulants. The
CC molecules have an integrin binding activity which when administered in
CC vivo results in the binding of the molecules to platelets thereby
CC inhibiting the aggregation of the platelets at sites of injury. Non-wild
CC type dendroaspin domains provide secondary, optionally further
CC functionality, e.g. antithrombotic action, inhibiting cell migration and
CC proliferation and regulating signal transduction. Such variants have bi-

CC or multifunctional activities against blood coagulation, particularly
CC thrombus formation and arterial/venous wall thickening at the sites of
CC injury. The variants may have activities against leukocyte recruitment,
CC immune system activation, tissue fibrosis and tumourigenesis. The
CC polypeptides can be used for the treatment or prophylaxis of a disease
CC associated with thrombosis, e.g. myocardial infarction, retinal
CC neovascularisation, endothelial injury, dysregulated apoptosis, abnormal
CC cell migration, leukocyte recruitment, immune system activation, tissue
CC fibrosis or tumorigenesis.
SQ Sequence 70 AA;

Query Match 44.9%; Score 35; DB 1; Length 70;
Best Local Similarity 22.2%; Pred. No. 63;
Matches 6; Conservative 6; Mismatches 3; Indels 12; Gaps 1;

Qy 1 RVCAAGRG----------ETCYRTV 15
|:| : || ::||: :
Db 1 RICFTPRGDMPGPYPGPCQEDSCYKNI 27

RESULT 11
W76684
ID W76684 standard; Protein; 72 AA.
AC W76684;
DT 05-JAN-1999 (first entry)
DE Elapidae modified dendroaspin protein fragment Den-Hr.
KW Dendroaspin; snake venom; clotting cascade; anticoagulant; platelet;
KW integrin binding; injury; blood; cell migration; thrombosis; inhibitor;
KW proliferation; signal transduction; regulator; coagulation; treatment;
KW prophylactic; artery; vein; wall thickening; myocardial infarction;
KW retinal neovascularisation; dysregulated apoptosis; tumorigenesis;
KW leukocyte recruitment, immune system; tissue fibrosis.
OS Elapidae.
OS Synthetic.
PN WO9842834-A1.
PD 01-OCT-1998.
PF 20-MAR-1998; G00848.
PR 20-MAR-1997; GB-005787.
PA (THRO-) THROMBOSIS RES INST.
PI Authi K, Kakkar V, Lu X, Scully MF;
DR WPI; 98-542278/46.
PT New hybrid dendroaspin polypeptide(s) - used for treating, e.g.
PT thrombosis, myocardial infarction, dysregulated apoptosis, abnormal
PT cell migration and immune system activation
PS Claim 6; Fig 3C; 59pp; English.
CC W76645-W76688 represent modified dendroaspin protein fragments isolated
CC from snake venom. When dendroaspin is modified to incorporate further
CC functional amino acid sequence, e.g. active portions or motifs of
CC agonists, antagonists or inhibitors of factors in the clotting cascade,
CC the resulting molecules are particularly useful as anticoagulants. The
CC molecules have an integrin binding activity which when administered in
CC vivo results in the binding of the molecules to platelets thereby
CC inhibiting the aggregation of the platelets at sites of injury. Non-wild
CC type dendroaspin domains provide secondary, optionally further
CC functionality, e.g. antithrombotic action, inhibiting cell migration and
CC proliferation and regulating signal transduction. Such variants have bi-
CC or multifunctional activities against blood coagulation, particularly
CC thrombus formation and arterial/venous wall thickening at the sites of
CC injury. The variants may have activities against leukocyte recruitment,
CC immune system activation, tissue fibrosis and tumourigenesis. The
CC polypeptides can be used for the treatment or prophylaxis of a disease
CC associated with thrombosis, e.g. myocardial infarction, retinal
CC neovascularisation, endothelial injury, dysregulated apoptosis, abnormal
CC cell migration, leukocyte recruitment, immune system activation, tissue
CC fibrosis or tumorigenesis.
SQ Sequence 72 AA;

Query Match 44.9%; Score 35; DB 1; Length 72;
Best Local Similarity 22.2%; Pred. No. 65;
Matches 6; Conservative 6; Mismatches 3; Indels 12; Gaps 1;

Qy 1 RVCAAGRG----------ETCYRTV 15
|:| : || ::||: :
Db 1 RICFTPRGDMPGPYPGPCQEDSCYKNI 27

RESULT 12
W76658
ID W76658 standard; Protein; 73 AA.
AC W76658;
DT 05-JAN-1999 (first entry)
DE Elapidae modified dendroaspin protein fragment DEN-HR11.
KW Dendroaspin; snake venom; clotting cascade; anticoagulant; platelet;
KW integrin binding; injury; blood; cell migration; thrombosis; inhibitor;
KW proliferation; signal transduction; regulator; coagulation; treatment;
KW prophylactic; artery; vein; wall thickening; myocardial infarction;
KW retinal neovascularisation; dysregulated apoptosis; tumorigenesis;
KW leukocyte recruitment, immune system; tissue fibrosis.
OS Elapidae.
OS Synthetic.
PN WO9842834-A1.
PD 01-OCT-1998.
PF 20-MAR-1998; G00848.
PR 20-MAR-1997; GB-005787.
PA (THRO-) THROMBOSIS RES INST.
PI Authi K, Kakkar V, Lu X, Scully MF;
DR WPI; 98-542278/46.
PT New hybrid dendroaspin polypeptide(s) - used for treating, e.g.
PT thrombosis, myocardial infarction, dysregulated apoptosis, abnormal
PT cell migration and immune system activation
PS Claim 6; Fig 3A; 59pp; English.
CC W76645-W76688 represent modified dendroaspin protein fragments isolated
CC from snake venom. When dendroaspin is modified to incorporate further
CC functional amino acid sequence, e.g. active portions or motifs of
CC agonists, antagonists or inhibitors of factors in the clotting cascade,
CC the resulting molecules are particularly useful as anticoagulants. The
CC molecules have an integrin binding activity which when administered in
CC vivo results in the binding of the molecules to platelets thereby
CC inhibiting the aggregation of the platelets at sites of injury. Non-wild
CC type dendroaspin domains provide secondary, optionally further
CC functionality, e.g. antithrombotic action, inhibiting cell migration and
CC proliferation and regulating signal transduction. Such variants have bi-
CC or multifunctional activities against blood coagulation, particularly
CC thrombus formation and arterial/venous wall thickening at the sites of
CC injury. The variants may have activities against leukocyte recruitment,
CC immune system activation, tissue fibrosis and tumourigenesis. The
CC polypeptides can be used for the treatment or prophylaxis of a disease
CC associated with thrombosis, e.g. myocardial infarction, retinal
CC neovascularisation, endothelial injury, dysregulated apoptosis, abnormal
CC cell migration, leukocyte recruitment, immune system activation, tissue
CC fibrosis or tumorigenesis.
SQ Sequence 73 AA;

Query Match 44.9%; Score 35; DB 1; Length 73;
Best Local Similarity 22.2%; Pred. No. 66;
Matches 6; Conservative 6; Mismatches 3; Indels 12; Gaps 1;

Qy 1 RVCAAGRG----------ETCYRTV 15
|:| : || ::||: :
Db 1 RICFTPRGDMPGPYPGPCQEDSCYKNI 27

RESULT 13
W76661
ID W76661 standard; Protein; 74 AA.
AC W76661;
DT 05-JAN-1999 (first entry)
DE Elapidae modified dendroaspin protein fragment DEN-TM11.
KW Dendroaspin; snake venom; clotting cascade; anticoagulant; platelet;
KW integrin binding; injury; blood; cell migration; thrombosis; inhibitor;
KW proliferation; signal transduction; regulator; coagulation; treatment;
KW prophylactic; artery; vein; wall thickening; myocardial infarction;
KW retinal neovascularisation; dysregulated apoptosis; tumorigenesis;

KW leukocyte recruitment, immune system; tissue fibrosis.
OS Elapidae.
OS Synthetic.
PN WO9842834-A1.
PD 01-OCT-1998.
PF 20-MAR-1998; G00848.
PR 20-MAR-1997; GB-005787.
PA (THRO-) THROMBOSIS RES INST.
PI Authi K, Kakkar V, Lu X, Scully MF;
DR WPI; 98-542278/46.
PT New hybrid dendroaspin polypeptide(s) - used for treating, e.g.
PT thrombosis, myocardial infarction, dysregulated apoptosis, abnormal
PT cell migration and immune system activation
PS Claim 6; Fig 3A; 59pp; English.
CC W76645-W76688 represent modified dendroaspin protein fragments isolated
CC from snake venom. When dendroaspin is modified to incorporate further
CC functional amino acid sequence, e.g. active portions or motifs of
CC agonists, antagonists or inhibitors of factors in the clotting cascade,
CC the resulting molecules are particularly useful as anticoagulants. The
CC molecules have an integrin binding activity which when administered in
CC vivo results in the binding of the molecules to platelets thereby
CC inhibiting the aggregation of the platelets at sites of injury. Non-wild
CC type dendroaspin domains provide secondary, optionally further
CC functionality, e.g. antithrombotic action, inhibiting cell migration and
CC proliferation and regulating signal transduction. Such variants have bi-
CC or multifunctional activities against blood coagulation, particularly
CC thrombus formation and arterial/venous wall thickening at the sites of
CC injury. The variants may have activities against leukocyte recruitment,
CC immune system activation, tissue fibrosis and tumourigenesis. The
CC polypeptides can be used for the treatment or prophylaxis of a disease
CC associated with thrombosis, e.g. myocardial infarction, retinal
CC neovascularisation, endothelial injury, dysregulated apoptosis, abnormal
CC cell migration, leukocyte recruitment, immune system activation, tissue
CC fibrosis or tumorigenesis.
SQ Sequence 74 AA;

Query Match 44.9%; Score 35; DB 1; Length 74;
Best Local Similarity 22.2%; Pred. No. 66;
Matches 6; Conservative 6; Mismatches 3; Indels 12; Gaps 1;

Qy 1 RVCAAGRG----------ETCYRTV 15
|:| : || ::||: :
Db 1 RICFTPRGDMPGPYPGPCQEDSCYKNI 27

RESULT 14
R89364
ID R89364 standard; peptide; 11 AA.
AC R89364;
DT 18-SEP-1996 (first entry)
DE MAGE-1 derived immunogenic peptide.
KW Immunogenic peptide; supermotif; HLA molecule; CTL response;
KW therapeutic; diagnostic; cancer; viral infection; hepatitis B;
KW hepatitis C.
OS Synthetic.
PN WO9603140-A1.
PD 08-FEB-1996.
PF 21-JUL-1995; U09234.
PR 21-JUL-1994; US-278634.
PR 23-NOV-1994; US-344824.
PR 30-MAY-1995; US-452843.
PA (CYTE-) CYTEL CORP.
PI Sette A, Sidney J;
DR WPI; 96-116784/12.
PT Compsn. comprising immunogenic peptide with supermotif allowing more
PT than one HLA mol. to bind - used to induce CTL response in patient
PT and for in vivo and ex vivo therapeutic and diagnostic applications
PS Claim 2; Page 26; 32pp; English.
CC The sequences given in R89362-82 are immunogenic peptides which were
CC use in the composition of the invention. The composition comprises
CC an immunogenic peptide of 9-10 residues with a supermotif which
CC allows binding of more than one HLA molecule. It pref. comprises
CC two conserved residues, a first at the 2nd position from the N-
CC terminal is Pro, and a 2nd at the C-terminal is Met. These peptides
CC are used to induce a CTL response in a patient. They are also
CC useful in compositions for in vivo and ex vivo therapeutic and
CC diagnostic applications, e.g the treatment of cancer and viral
CC infections, e.g. hepatitis B and C.
SQ Sequence 11 AA;

Query Match 38.5%; Score 30; DB 1; Length 11;
Best Local Similarity 57.1%; Pred. No. 68;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 9 ETCYRTV 15
|:|:|:|
Db 4 ESCFRAV 10

RESULT 15
W03034
ID W03034 standard; peptide; 24 AA.
AC W03034;
DT 17-FEB-1997 (first entry)
DE Thrombolytic enzyme fragment #3.
KW Thrombolytic enzyme; destabilase; hydrolysis; D fragment dimer;
KW epsilon-(gamma-glutamyl)-lysine isopeptide bond; stabilised fibrin;
KW salivary gland; medicinal leech.
OS Hirudo medicinalis.
PN WO9619999-A1.
PD 04-JUL-1996.
PF 25-MAY-1995; RU0102.
PR 23-DEC-1994; RU-043698.
PA (BIOO-) INST BIOORG KHIM IM M M SHEMY.
PI Barsova EV, Baskova IP, Bogdanova EA, Lukyanov SA;
PI Sverdlov ED, Zavalova LL;
DR WPI; 96-321638/32.
PT Thrombolytic enzyme destabilase - selectively hydrolysing stabilised
PT fibrin
PS Disclosure; Page 3; 16pp; Russian
CC The sequences given in W03032-35 represent CNBr fragments of the novel
CC thrombolytic enzyme of the invention. The thrombolytic enzyme, also
CC known as destabilase, is capable of hydrolysing exo- and endo-
CC epsilon-(gamma-glutamyl)-lysine isopeptide bonds in synthetic
CC substrates and in stabilised fibrin and its D fragment dimer.
CC Destabilase may be used as a thrombolytic agent in clinical and
CC experimental medicine. It can be isolated from the salivary glands
CC of medicinal leeches.
SQ Sequence 24 AA;

Query Match 39.7%; Score 31; DB 1; Length 24;
Best Local Similarity 55.6%; Pred. No. 96;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 CAAGRGETC 11
|::|| ||
Db 7 CTGGRQPTC 15

Search completed: May 4, 1999, 12:32:41
Job time: 9788 sec

OM protein - protein search, using sw model

Run on: May 4, 1999, 08:18:15 ; Search time 25.07 Seconds
(without alignments)
22.413 Million cell updates/sec

Title: US-09-037-460-2_COPY_30_44
Perfect score: 74
Sequence: 1 QHCDSSECKSSPRCK 15

Scoring table: PAM150

Searched: 116738 seqs, 37460341 residues

Database : PIR_58:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	40	54.1	57	2	S59073	metallothionein is
2	42	56.8	157	2	S39849	pfrA protein - Pse
3	44	59.5	861	2	A48825	Notch homolog Motc
4	42	56.8	488	1	EXHU	coagulation factor
5	38	51.4	75	2	S17156	metallothionein -
6	40	54.1	251	2	A55035	cysteine-rich prot
7	37	50.0	61	2	S14049	metallothionein pr
8	36	48.6	39	2	A46057	thrombin inhibitor
9	41	55.4	473	2	A56175	adhesive plaque pr
10	42	56.8	772	2	S32659	integrin beta 2 ch
11	36	48.6	57	1	SMKD2S	metallothionein 2
12	44	59.5	2524	2	A35844	Xotch protein - Af
13	44	59.5	2531	2	A46019	gene Notch-1 prote
14	37	50.0	111	2	S44787	D2007.1 protein -
15	36	48.6	77	2	S36032	thrombin inhibitor
16	39	52.7	400	2	S32804	beta-3-adrenergic
17	39	52.7	400	2	A41679	beta-3-adrenergic
18	39	52.7	400	2	A53281	beta 3-adrenergic
19	37.5	50.7	200	2	S35292	hypothetical prote
20	39	52.7	410	2	S15163	probable transposa
21	37	50.0	180	2	A45810	glycoprotein antig
22	38	51.4	319	2	A53502	follistatin - Afri
23	35	47.3	79	2	H69193	2-oxoisovalerate o
24	38	51.4	335	2	S71796	centrosome-binding
25	38	51.4	343	2	S55369	follistatin - chic
26	38	51.4	379	1	DEILSP	alcohol dehydrogen
27	38	51.4	379	1	S01893	alcohol dehydrogen
28	42	56.8	2555	2	A40043	notch protein homo
29	36	48.6	155	2	A45293	conopressin precur
30	39.5	53.4	837	2	A42112	mucin-like peptide
31	38	51.4	414	2	S36838	acetyl-CoA C-acylt
32	39	52.7	683	2	JC5393	zinc finger protei
33	38	51.4	466	1	KFHU7	coagulation factor
34	34	45.9	75	2	B45206	metallothionein 2
35	35	47.3	125	1	NFRT1	oxytocin / neuroph
36	35	47.3	125	2	A43755	oxytocin / neuroph
37	34	45.9	78	2	S09414	proteinase inhibit
38	40	54.1	1375	2	F48216	neurexin III-alpha
39	40	54.1	1378	2	E48216	neurexin III-alpha
40	40	54.1	1392	2	D48216	neurexin III-alpha
41	37	50.0	337	2	I47079	follistatin - shee
42	40	54.1	1395	2	C48216	neurexin III-alpha
43	40	54.1	1396	2	H48216	neurexin III-alpha
44	40	54.1	1399	2	G48216	neurexin III-alpha
45	37	50.0	343	2	S45321	follistatin - mous

ALIGNMENTS

RESULT 1
S59073
metallothionein isoform IIa - blue crab
C;Species: Callinectes sapidus
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 29-Aug-1997
C;Accession: S59073
R;Brouwer, M.; Enghild, J.; Hoexum-Brouwer, T.; Thogersen, I.; Truncali, A.
Biochem. J. 311, 617-622, 1995
A;Title: Primary structure and tissue-specific expression of blue crab (Callinectes s
A;Reference number: S59072
A;Accession: S59073
A;Molecule type: protein
A;Residues: 1-57 <BRO>
C;Superfamily: metallothionein
C;Keywords: metal binding

Query Match 54.1%; Score 40; DB 2; Length 57;
Best Local Similarity 46.2%; Pred. No. 7;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```
Qy       3 CDSSECKSSPRCK 15
           |  :|||:: :||
Db      11 CKEGECKTGCKCK 23
```

RESULT 2
S39849
pfrA protein - Pseudomonas putida
C;Species: Pseudomonas putida
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 01-Aug-1997
C;Accession: S39849
R;Venturi, V.; Ottevanger, C.; Leong, J.; Weisbeek, P.J.
Mol. Microbiol. 10, 63-73, 1993
A;Title: Identification and characterization of a siderophore regulatory gene (pfrA)
A;Reference number: S39849
A;Accession: S39849
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-157 <VEN>

Query Match 56.8%; Score 42; DB 2; Length 157;
Best Local Similarity 46.2%; Pred. No. 7.3;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 QHCDSSECKSSPR 13
           :||: :||| : |
Db     107 DHCEKGECKDPER 119
```

RESULT 3
A48825
Notch homolog Motch protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 14-Aug-1998
C;Accession: A48825
R;Reaume, A.G.; Conlon, R.A.; Zirngibl, R.; Yamaguchi, T.P.; Rossant, J.
Dev. Biol. 154, 377-387, 1992
A;Title: Expression analysis of a Notch homologue in the mouse embryo.
A;Reference number: A48825; MUID:93050801
A;Accession: A48825

A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-861 <REA>
A;Experimental source: embryo
A;Note: sequence extracted from NCBI backbone (NCBIP:119144)
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
F;26-57/Domain: EGF homology <EGF>

Query Match 59.5%; Score 44; DB 2; Length 861;
Best Local Similarity 58.3%; Pred. No. 12;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 3 CDSSECKSSPRC 14
|||: ||:: ||
Db 274 CDSAPCKNGGRC 285

RESULT 4
EXHU
coagulation factor Xa (EC 3.4.21.6) precursor - human
N;Alternate names: Stuart factor
C;Species: Homo sapiens (man)
C;Date: 15-Nov-1984 #sequence_revision 02-May-1994 #text_change 24-Oct-1997
C;Accession: A24478; JQ0917; A42485; A25853; A22208; A21284; A20362; S39415; I54051; A00
R;Leytus, S.P.; Foster, D.C.; Kurachi, K.; Davie, E.W.
Biochemistry 25, 5098-5102, 1986
A;Title: Gene for human Factor X: a blood coagulation factor whose gene organization is
A;Reference number: A24478; MUID:87026600
A;Accession: A24478
A;Molecule type: DNA
A;Residues: 1-488 <LEY>
A;Cross-references: GB:L29433; GB:M14327; NID:g459809; PID:g182831
R;Messier, T.L.; Pittman, D.D.; Long, G.L.; Kaufman, R.J.; Church, W.R.
Gene 99, 291-294, 1991
A;Title: Cloning and expression in COS-1 cells of a full-length cDNA encoding human coag
A;Reference number: JQ0917; MUID:91216473
A;Accession: JQ0917
A;Molecule type: mRNA
A;Residues: 1-488 <MES>
A;Cross-references: GB:M57285; NID:g182389; PID:g182390
R;Miao, C.H.; Leytus, S.P.; Chung, D.W.; Davie, E.W.
J. Biol. Chem. 267, 7395-7401, 1992
A;Title: Liver-specific expression of the gene coding for human factor X, a blood coagul
A;Reference number: A42485; MUID:92218390
A;Accession: A42485
A;Molecule type: DNA
A;Residues: 1-15 <MIA>
A;Experimental source: liver
A;Note: sequence extracted from NCBI backbone (NCBIN:93780, NCBIP:93787)
R;Kaul, R.K.; Hildebrand, B.; Roberts, S.; Jagadeeswaran, P.
Gene 41, 311-314, 1986
A;Title: Isolation and characterization of human blood-coagulation factor X cDNA.
A;Reference number: A25853; MUID:86221713
A;Accession: A25853
A;Molecule type: mRNA
A;Residues: 19-284,'E',289-488 <KAU>
A;Cross-references: GB:M22613; NID:g180335; PID:g180336
R;Fung, M.R.; Hay, C.W.; MacGillivray, R.T.A.
Proc. Natl. Acad. Sci. U.S.A. 82, 3591-3595, 1985
A;Title: Characterization of an almost full-length cDNA coding for human blood coagulati
A;Reference number: A22208; MUID:85216545
A;Accession: A22208
A;Molecule type: mRNA
A;Residues: 13-441,'S',443-488 <FUN>
A;Cross-references: GB:K03194; NID:g182840; PID:g182841
R;Leytus, S.P.; Chung, D.W.; Kisiel, W.; Kurachi, K.; Davie, E.W.
Proc. Natl. Acad. Sci. U.S.A. 81, 3699-3702, 1984
A;Title: Characterization of a cDNA coding for human factor X.
A;Reference number: A21284; MUID:84222026
A;Accession: A21284
A;Molecule type: mRNA
A;Residues: 13-284,'E',289-488 <LE2>
A;Cross-references: GB:K01886
R;McMullen, B.A.; Fujikawa, K.; Kisiel, W.; Sasagawa, T.; Howald, W.N.; Kwa, E.Y.; We
Biochemistry 22, 2875-2884, 1983
A;Title: Complete amino acid sequence of the light chain of human blood coagulation f
A;Reference number: A20362; MUID:83257207
A;Accession: A20362
A;Molecule type: protein
A;Residues: 41-179 <MCM>
R;Inoue, K.; Morita, T.
Eur. J. Biochem. 218, 153-163, 1993
A;Title: Identification of O-linked oligosaccharide chains in the activation peptides
A;Reference number: S39414
A;Accession: S39415
A;Molecule type: protein
A;Residues: 183-234 <INO>
A;Note: glycosylation sites
A;Note: identification and characterization of beta-hydroxyaspartic acid
R;Jagadeeswaran, P.; Reddy, S.V.; Rao, K.J.; Hamsabhushanam, K.; Lyman, G.
Gene 84, 517-519, 1989
A;Title: Cloning and characterization of the 5' end (exon 1) of the gene encoding hum
A;Reference number: I54051; MUID:90128299
A;Accession: I54051
A;Status: translation not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-23 <RES>
A;Cross-references: GB:M33297; NID:g183860; PID:g553330
R;Padmanabhan, K.; Padmanabhan, K.P.; Tulinsky, A.; Park, C.H.; Bode, W.; Huber, R.;
J. Mol. Biol. 232, 947-966, 1993
A;Title: Structure of human des(1-45) factor Xa at 2.2 angstroms resolution.
A;Reference number: A49458
A;Contents: annotation; X-ray crystallography, 2.2 angstroms
C;Comment: The two chains held together by one disulfide bond are formed from a singl
C;Comment: The activation peptide is cleaved by factor IXa (in the intrinsic pathway)
C;Genetics:
A;Gene: GDB:F10
A;Cross-references: GDB:119890; OMIM:227600
A;Map position: 13q34-13q34
A;Introns: 24/1; 77/3; 86/1; 124/1; 150/3; 249/3; 289/1
A;Note: deficiency of this factor causes Stuart disease
C;Function:
A;Description: catalyzes the proteolytic activation of prothrombin to thrombin in the
A;Pathway: blood coagulation
C;Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homol
C;Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglu
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-40/Domain: propeptide #status predicted <PRO>
F;25-84/Domain: Gla domain homology <GLA>
F;41-179/Product: coagulation factor X light chain #status experimental <LCH>
F;90-121/Domain: EGF homology <EGF1>
F;129-164/Domain: EGF homology <EGF2>
F;183-488/Product: coagulation factor X heavy chain #status experimental <HCH>
F;183-234/Domain: activation peptide #status experimental <APT>
F;235-488/Product: coagulation factor Xa heavy chain #status experimental <ACT>
F;235-462/Domain: trypsin homology <TRY>
F;46,47,54,56,59,60,65,66,69,72,79/Modified site: gamma-carboxyglutamic acid (Glu) #s
F;57-62/Disulfide bonds: #status predicted
F;90-101,95-110,112-121,129-140,136-149,151-164,172-342,241-246,261-277,390-404,415-4
F;103/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status experimental
F;199,211/Binding site: carbohydrate (Thr) (covalent) #status experimental
F;221,231/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;234-235/Cleavage site: Arg-Ile (coagulation factor IXa, coagulation factor VIIa) #s
F;276,322,419/Active site: His, Asp, Ser #status experimental

Query Match 56.8%; Score 42; DB 1; Length 488;
Best Local Similarity 33.3%; Pred. No. 16;
Matches 5; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QHCDSSECKSSPRCK 15
::|::| | : :||
Db 88 DQCETSPCQNQGKCK 102

```
RESULT   5
S17156
metallothionein - eastern oyster
C;Species: Crassostrea virginica (eastern oyster)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Sep-1997
C;Accession: S17156
R;Unger, M.E.; Chen, T.T.; Murphy, C.M.; Vestling, M.M.; Fenselau, C.; Roesijadi, G.
Biochim. Biophys. Acta 1074, 371-377, 1991
A;Title: Primary structure of molluscan metallothioneins deduced from PCR-amplified cDNA
A;Reference number: S17156; MUID:91363394
A;Accession: S17156
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-75 <UNG>
A;Cross-references: EMBL:X59862; NID:g288277; PID:g288278
C;Superfamily: metallothionein


  Query Match             51.4%;  Score 38;  DB 2;  Length 75;
  Best Local Similarity   46.2%;  Pred. No. 16;
  Matches    6;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       3 CDSSECKSSPRCK 15
           | :: || :| ||
Db      19 CPATGCKCGPGCK 31


RESULT   6
A55035
cysteine-rich protein CRP1 - earthworm (Enchytraeus buchholzi)
C;Species: Enchytraeus buchholzi
C;Date: 14-Nov-1994 #sequence_revision 03-Nov-1995 #text_change 10-Sep-1997
C;Accession: A55035; S45034
R;Willuhn, J.; Schmitt-Wrede, H.P.; Greven, H.; Wunderlich, F.
J. Biol. Chem. 269, 24688-24691, 1994
A;Title: cDNA cloning of a cadmium-inducible mRNA encoding a novel cysteine-rich, non-me
A;Reference number: A55035
A;Accession: A55035
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-251 <WIL>
A;Cross-references: EMBL:X79344; NID:g488802; PID:g488803


  Query Match             54.1%;  Score 40;  DB 2;  Length 251;
  Best Local Similarity   38.5%;  Pred. No. 20;
  Matches    5;  Conservative    5;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       3 CDSSECKSSPRCK 15
           ||: :|| :: |:
Db     223 CDNVNCKCGSSCR 235


RESULT   7
S14049
metallothionein precursor - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YHR053c; protein YHR055c
C;Species: Saccharomyces cerevisiae
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 05-Dec-1997
C;Accession: S14049; A29373; A17610; A24123; S46703; S46705
R;Jeyaprakash, A.; Welch, J.W.; Fogel, S.
Mol. Gen. Genet. 225, 363-368, 1991
A;Title: Multicopy CUP1 plasmids enhance cadmium and copper resistance levels in yeast.
A;Reference number: S14049; MUID:91203809
A;Accession: S14049
A;Molecule type: DNA
A;Residues: 1-61 <MOL>
R;Karin, M.; Najarian, R.; Haslinger, A.; Valenzuela, P.; Welch, J.; Fogel, S.
Proc. Natl. Acad. Sci. U.S.A. 81, 337-341, 1984
A;Title: Primary structure and transcription of an amplified genetic locus: the CUP1 loc
A;Reference number: A29373; MUID:84119482
A;Accession: A29373
A;Molecule type: DNA
A;Residues: 1-61 <KAR>
A;Cross-references: EMBL:K02204; NID:g171337; PID:g171338
R;Butt, T.R.; Sternberg, E.J.; Gorman, J.A.; Clark, P.; Hamer, D.; Rosenberg, M.; Cro
Proc. Natl. Acad. Sci. U.S.A. 81, 3332-3336, 1984
A;Title: Copper metallothionein of yeast, structure of the gene, and regulation of ex
A;Reference number: A17610; MUID:84221953
A;Accession: A17610
A;Molecule type: DNA
A;Residues: 1-61 <BUT>
A;Cross-references: EMBL:K02204; NID:g171337; PID:g171338
R;Winge, D.R.; Nielson, K.B.; Gray, W.R.; Hamer, D.H.
J. Biol. Chem. 260, 14464-14470, 1985
A;Title: Yeast metallothionein. Sequence and metal-binding properties.
A;Reference number: A92506; MUID:86033949
A;Accession: A24123
A;Molecule type: protein
A;Residues: 9-61 <WIN>
R;Latreille, P.
submitted to the EMBL Data Library, May 1994
A;Description: The sequence of S. cerevisiae cosmid 8025.
A;Reference number: S46691
A;Accession: S46703
A;Molecule type: DNA
A;Residues: 1-61 <LAT>
A;Cross-references: EMBL:U00061; NID:g487943; PID:e108807; MIPS:YHR053c
A;Genetics: CUP1A
A;Accession: S46705
A;Molecule type: DNA
A;Residues: 1-61 <LA2>
A;Cross-references: EMBL:U00061; NID:g487943; PID:g487953; MIPS:YHR055c
A;Genetics: CUP1B
C;Genetics: <CUP1A>
A;Gene: SGD:CUP1; CUP1A
A;Cross-references: MIPS:YHR053c; SGD:S0001095
A;Map position: 8R
C;Genetics: <CUP1B>
A;Gene: SGD:CUP1; CUP1B
A;Cross-references: MIPS:YHR055c; SGD:S0001097
A;Map position: 8R
C;Superfamily: metallothionein


  Query Match             50.0%;  Score 37;  DB 2;  Length 61;
  Best Local Similarity   30.8%;  Pred. No. 20;
  Matches    4;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 HCDSSECKSSPRC 14
           :|: : ||:: :|
Db      16 QCQCGSCKNNEQC 28


RESULT   8
A46057
thrombin inhibitor hemadin - terrestrial leech (Haemadipsa sylvestris) (fragment)
C;Species: Haemadipsa sylvestris
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
C;Accession: A46057
R;Strube, K.H.; Kroger, B.; Bialojan, S.; Otte, M.; Dodt, J.
J. Biol. Chem. 268, 8590-8595, 1993
A;Title: Isolation, sequence analysis, and cloning of haemadin. An anticoagulant pept
A;Reference number: A46057; MUID:93232009
A;Accession: A46057
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-39 <STR>
A;Cross-references: GB:S58792; NID:g302486; PID:g302487
A;Note: sequence extracted from NCBI backbone (NCBIN:129610, NCBIP:129611)


  Query Match             48.6%;  Score 36;  DB 2;  Length 39;
  Best Local Similarity   30.8%;  Pred. No. 20;
  Matches    4;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;
```

```
Qy      1 QHCDSSECKSSPR 13
          | |: ::| : |:
Db     12 QSCNDGQCSGDPK 24
```

RESULT 9
A56175
adhesive plaque protein Mgfp2 precursor - Mediterranean mussel
C;Species: Mytilus galloprovincialis (Mediterranean mussel)
C;Date: 27-Apr-1995 #sequence_revision 03-Oct-1995 #text_change 10-Sep-1997
C;Accession: A56175
R;Inoue, K.; Takeuchi, Y.; Miki, D.; Odo, S.
J. Biol. Chem. 270, 6698-6701, 1995
A;Title: Mussel adhesive plaque protein gene is a novel member of epidermal growth facto
A;Reference number: A56175
A;Accession: A56175
A;Molecule type: mRNA
A;Residues: 1-473 <INO>
A;Cross-references: GB:D43794; NID:g602767; PID:d1008438; PID:g602768
C;Keywords: duplication
F;1-17/Domain: signal sequence #status predicted <SIG>
F;23,36,43,56,75,382,424,455,468,473/Modified site: 3',4'-dihydroxyphenylalanine (Tyr) #

```
  Query Match             55.4%;  Score 41;  DB 2;  Length 473;
  Best Local Similarity   41.7%;  Pred. No. 22;
  Matches    5;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 CDSSECKSSPRC 14
          | :::||: :||
Db     86 CKPNQCKNKSRC 97
```

RESULT 10
S32659
integrin beta 2 chain (CD18) - chicken
N;Alternate names: CD18 protein
C;Species: Gallus gallus (chicken)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Nov-1996
C;Accession: I50660; S32659
R;Bilsland, C.A.; Springer, T.A.
J. Leukoc. Biol. 55, 501-506, 1994
A;Title: Cloning and expression of the chicken CD18 cDNA.
A;Reference number: I50660; MUID:94194252
A;Accession: I50660
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-772 <BI2>
A;Cross-references: EMBL:X71786; NID:g297566; PID:g297567
C;Superfamily: integrin beta chain

```
  Query Match             56.8%;  Score 42;  DB 2;  Length 772;
  Best Local Similarity   50.0%;  Pred. No. 22;
  Matches    6;  Conservative    4;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 HCDSSECKSSPR 13
          :|| :||| :|:
Db    557 RCDCGECKCTPK 568
```

RESULT 11
SMKD2S
metallothionein 2 - mud crab
C;Species: Scylla serrata (mud crab)
C;Date: 19-Feb-1984 #sequence_revision 19-Feb-1984 #text_change 13-Sep-1996
C;Accession: A03284
R;Lerch, K.; Ammer, D.; Olafson, R.W.
J. Biol. Chem. 257, 2420-2426, 1982
A;Title: Crab metallothionein. Primary structures of metallothioneins 1 and 2.
A;Reference number: A92363; MUID:82142340
A;Accession: A03284
A;Molecule type: protein
A;Residues: 1-57 <LER>
C;Superfamily: metallothionein
C;Keywords: metal binding

```
  Query Match             48.6%;  Score 36;  DB 1;  Length 57;
  Best Local Similarity   41.7%;  Pred. No. 26;
  Matches    5;  Conservative    4;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 CDSSECKSSPRC 14
          |  :|||:: :|
Db     11 CKEGECKTGCKC 22
```

RESULT 12
A35844
Xotch protein - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 12-Oct-1990 #sequence_revision 12-Oct-1990 #text_change 14-Aug-1998
C;Accession: A35844
R;Coffman, C.; Harris, W.; Kintner, C.
Science 249, 1438-1441, 1990
A;Title: Xotch, the Xenopus homolog of Drosophila notch.
A;Reference number: A35844; MUID:90385285
A;Accession: A35844
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A;Molecule type: mRNA
A;Residues: 1-2524 <COF>
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol
C;Keywords: transmembrane protein
F;222-254/Domain: EGF homology <EGF>
F;1924-1956/Domain: ankyrin repeat homology <AN1>
F;1957-1989/Domain: ankyrin repeat homology <AN2>
F;1991-2023/Domain: ankyrin repeat homology <AN3>
F;2024-2056/Domain: ankyrin repeat homology <AN4>
F;2057-2089/Domain: ankyrin repeat homology <AN5>

```
  Query Match             59.5%;  Score 44;  DB 2;  Length 2524;
  Best Local Similarity   53.8%;  Pred. No. 27;
  Matches    7;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      3 CDSSECKSSPRCK 15
          | :| ||:: |||
Db    833 CAGSPCKNGGRCK 845
```

RESULT 13
A46019
gene Notch-1 protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 14-Aug-1998
C;Accession: A46019
R;del Amo, F.F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; G
Genomics 15, 259-264, 1993
A;Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog
A;Reference number: A46019; MUID:93194170
A;Accession: A46019
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-2531 <DEL>
A;Cross-references: GB:Z11886; GB:S47228; NID:g288502; PID:g288503
A;Note: sequence extracted from NCBI backbone (NCBIP:127318)
C;Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol
F;757-788/Domain: EGF homology <EGF>
F;1917-1948/Domain: ankyrin repeat homology <AN1>
F;1949-1981/Domain: ankyrin repeat homology <AN2>
F;1983-2015/Domain: ankyrin repeat homology <AN3>
F;2016-2048/Domain: ankyrin repeat homology <AN4>
F;2049-2081/Domain: ankyrin repeat homology <AN5>

```
  Query Match             59.5%;  Score 44;  DB 2;  Length 2531;
```

Best Local Similarity 58.3%; Pred. No. 27;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```
Qy       3 CDSSECKSSPRC 14
           ||| : ||:: ||
Db    1063 CDSAPCKNGGRC 1074
```

RESULT 14
S44787
D2007.1 protein - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 09-Sep-1997
C;Accession: S44787
R;Favello, A.D.
submitted to the EMBL Data Library, May 1993
A;Description: Sequence of the C. elegans cosmid D2007.
A;Reference number: S44617
A;Accession: S44787
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-111 <FAV>
A;Cross-references: EMBL:L16560; NID:g289666; PID:g289668
C;Genetics:
A;Introns: 48/1; 84/3

Query Match 50.0%; Score 37; DB 2; Length 111;
Best Local Similarity 33.3%; Pred. No. 30;
Matches 5; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 QHCDSSECKSSPRCK 15
           ::| :: ||:| :|:
Db      91 DQCGNACCKTSEQCR 105
```

RESULT 15
S36032
thrombin inhibitor haemadin - terrestrial leech (Haemadipsa sylvestris)
C;Species: Haemadipsa sylvestris
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Sep-1997
C;Accession: S36032; B46057
R;Krwger, B.
submitted to the EMBL Data Library, January 1993
A;Reference number: S36032
A;Accession: S36032
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-77 <KRV>
A;Cross-references: EMBL:Z19864; NID:g298108; PID:g298109
R;Strube, K.H.; Kroger, B.; Bialojan, S.; Otte, M.; Dodt, J.
J. Biol. Chem. 268, 8590-8595, 1993
A;Title: Isolation, sequence analysis, and cloning of haemadin. An anticoagulant peptide
A;Reference number: A46057; MUID:93232009
A;Accession: B46057
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-77 <STR>
A;Note: sequence extracted from NCBI backbone (NCBIN:129612, NCBIP:129614)

Query Match 48.6%; Score 36; DB 2; Length 77;
Best Local Similarity 30.8%; Pred. No. 33;
Matches 4; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```
Qy       1 QHCDSSECKSSPR 13
           | |: ::| : |:
Db      50 QSCNDGQCSGDPK 62
```

Search completed: May 4, 1999, 08:18:16
Job time: 6968 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 4, 1999, 12:32:37 ; Search time 27.23 Seconds
(without alignments)
11.141 Million cell updates/sec

Title: US-09-037-460-2_COPY_30_44
Perfect score: 74
Sequence: 1 QHCDSSECKSSPRCK 15

Scoring table: PAM150

Searched: 162890 seqs, 20225328 residues

Database : A_Geneseq_34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	74	100.0	184	1	R98994	Vascular IBP-like
2	39.5	53.4	50	1	R96122	Leech derived fahs
3	39.5	53.4	50	1	R96123	Leech derived fahs
4	42	56.8	139	1	R95596	Factor X light cha
5	44	59.5	467	1	W40283	Human Factor X pro
6	42	56.8	250	1	R13675	Factor X-LACI hybr
7	43	58.1	400	1	R32501	Beta-adrenergic re
8	37	50.0	52	1	W80576	Human factor VII E
9	37	50.0	53	1	R40210	Sequence of yeast
10	36	48.6	39	1	R37437	Thrombin inhibitor
11	37	50.0	61	1	P94625	CUP1 translation p
12	37	50.0	61	1	P50776	Sequence encoded b
13	37	50.0	61	1	P30006	Sequence of yeast
14	37	50.0	61	1	R54087	Metallothionein. P
15	42	56.8	436	1	R22513	Truncated precurso
16	42	56.8	437	1	R88468	Human two chain fa
17	42	56.8	448	1	R37402	Factor X. Factor X
18	42	56.8	448	1	R35762	Factor X (X). Seri
19	42	56.8	448	1	W66092	Human factor X var
20	34	45.9	21	1	W00600	Heparan sulphate p
21	44	59.5	1055	1	W44298	Human serrate 2 pr
22	42	56.8	488	1	R22512	Mutated precursor
23	42	56.8	488	1	R22511	Human Factor Xai.
24	42	56.8	488	1	W76219	Human Factor X pro
25	42	56.8	488	1	W76216	Human Factor X pro
26	42	56.8	488	1	W76217	Human Factor X pro
27	42	56.8	488	1	W76218	Human Factor X pro
28	37	50.0	70	1	P94628	CUP1 translation p
29	37	50.0	70	1	P94627	CUP1 translation p
30	44	59.5	1212	1	W44299	Human serrate 2. H
31	44	59.5	1257	1	W05834	Human Serrate-2 (H
32	37	50.0	85	1	R62621	Deduced sequence o
33	37	50.0	86	1	R62620	Recombinant factor
34	41	55.4	473	1	R86869	Adhesive protein.
35	37	50.0	110	1	R62622	Deduced sequence e
36	36	48.6	77	1	R37438	Thrombin inhibitor
37	37	50.0	148	1	W68544	Protein encoded by
38	37	50.0	153	1	R62516	Deduced sequence e
39	37	50.0	153	1	R62624	Deduced sequence o
40	39	52.7	388	1	R25698	Murine adrenergic
41	37	50.0	182	1	R62619	Deduced sequence o
42	30	40.5	12	1	R64297	Factor-VII-derived
43	30	40.5	12	1	R67863	FVII/TF multi-prot
44	39	52.7	400	1	R54992	Murine beta-3 adre
45	33	44.6	44	1	R96225	Novel growth facto

ALIGNMENTS

RESULT 1
R98994
ID R98994 standard; Protein; 184 AA.
AC R98994;
DT 06-NOV-1996 (first entry)
DE Vascular IBP-like growth factor.
KW Vascular IBF-like growth factor; VIGF;
KW insulin-like growth factor binding protein; agonist; antagonist;
KW muscle wastage; osteoporosis; implant fixation; wound healing;
KW therapy; diagnosis.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1. .21
FT /label= Sig_peptide
PN WO9617931-A1.
PD 13-JUN-1996.
PF 09-DEC-1994; U14388.
PR 09-DEC-1994; WO-U14388.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Hastings GA, Rosen CA;
DR WPI; 96-287176/29.
DR N-PSDB; T34991.
PT Human vascular insulin-like growth factor binding protein-like
PT growth factor, and its nucleic acid sequence and (ant)agonists -
PT used, e.g. to treat muscle wasting diseases or aid implant fixation,
PT or limit excess connective tissue prodn. during wound healing.
PS Claim 14; Page 43-44; 61pp; English.
CC Human vascular insulin-like growth factor binding protein-like
CC growth factor (R98994), or VIGF, is a protein of primarily
CC vascular origin that is structurally related to the IBP and CCN
CC protein families. It can be expressed in e.g. E. coli, CHO or
CC insect host cells using a vector incorporating a cDNA clone
CC (T34991), or its derivative, obtd. from human umbilical
CC endothelial cells. It is useful therapeutically e.g. for
CC treating muscle wasting diseases or osteoporosis, or can be used
CC to detect diseases associated with under- or over-expression of VIGF,
CC or to screen for antagonists useful during wound healing.
SQ Sequence 184 AA;

Query Match 100.0%; Score 74; DB 1; Length 184;
Best Local Similarity 100.0%; Pred. No. 0.00048;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QHCDSSECKSSPRCK 15
||||||||||||||||
Db 30 QHCDSSECKSSPRCK 44

RESULT 2
R96122
ID R96122 standard; Peptide; 50 AA.
AC R96122;
DT 17-DEC-1996 (first entry)
DE Leech derived fahsin based protease inhibitor #2.
KW Protease inhibitor; isoform; elastase; chymotrypsin; trypsin; leech;
KW tissue; secretion; saliva; fahsin; antibiotic; diabetes mellitus;
KW blood clotting disorder; neutrophil function; emphysema;
KW rheumatoid arthritis; HIV infection; human immunodeficiency virus.
OS Limnatis nilotica.
PN WO9613585-A1.
PD 09-MAY-1996.
PF 27-OCT-1995; E04223.
PR 28-OCT-1994; EP-117053.
PR 14-MAR-1995; EP-103637.
PA (CLOD-) CLODICA SA.

```
PI   Voerman G;
DR   WPI; 96-239498/24.
PT   New protease inhibitors from the leech Limnatis nilotica - for
PT   treating, e.g. blood clotting disorders, HIV infection, diabetes
PT   mellitus etc.
PS   Claim 3; Page 26; 41pp; English.
CC   The protease inhibitor peptide isoforms given in R96121-23 are
CC   elastase/chymotrypsin- and trypsin inhibitors which may be isolated
CC   from leech tissue or leech secretions, e.g. saliva. These peptides
CC   belong to the family of leech derived substances named fahsin's which
CC   also have an antibiotic effect. The fahsin family of proteins comprise
CC   50/51 amino acids and occur in various isoforms. These peptides are
CC   useful in the treatment of diabetes mellitus, blood clotting disorders,
CC   disorders of neutrophil function, e.g. emphysema, rheumatoid arthritis,
CC   HIV infection and other immunological and inflammatory diseases.
SQ   Sequence   50 AA;


  Query Match             53.4%;  Score 39.5;  DB 1;  Length 50;
  Best Local Similarity   43.8%;  Pred. No. 11;
  Matches    7;  Conservative    4;  Mismatches    4;  Indels    1;  Gaps    1;

Qy       1 QHCDSSECKSSP-RCK 15
           | |  ::|| :| ||:
Db      13 QLCVDGQCKCTPIRCR 28


RESULT   3
R96123
ID   R96123 standard; Peptide; 50 AA.
AC   R96123;
DT   17-DEC-1996  (first entry)
DE   Leech derived fahsin based protease inhibitor #3.
KW   Protease inhibitor; isoform; elastase; chymotrypsin; trypsin; leech;
KW   tissue; secretion; saliva; fahsin; antibiotic; diabetes mellitus;
KW   blood clotting disorder; neutrophil function; emphysema;
KW   rheumatoid arthritis; HIV infection; human immunodeficiency virus.
OS   Limnatis nilotica.
PN   WO9613585-A1.
PD   09-MAY-1996.
PF   27-OCT-1995; E04223.
PR   28-OCT-1994; EP-117053.
PR   14-MAR-1995; EP-103637.
PA   (CLOD-) CLODICA SA.
PI   Voerman G;
DR   WPI; 96-239498/24.
PT   New protease inhibitors from the leech Limnatis nilotica - for
PT   treating, e.g. blood clotting disorders, HIV infection, diabetes
PT   mellitus etc.
PS   Claim 3; Page 26; 41pp; English.
CC   The protease inhibitor peptide isoforms given in R96121-23 are
CC   elastase/chymotrypsin- and trypsin inhibitors which may be isolated
CC   from leech tissue or leech secretions, e.g. saliva. These peptides
CC   belong to the family of leech derived substances named fahsin's which
CC   also have an antibiotic effect. The fahsin family of proteins comprise
CC   50/51 amino acids and occur in various isoforms. These peptides are
CC   useful in the treatment of diabetes mellitus, blood clotting disorders,
CC   disorders of neutrophil function, e.g. emphysema, rheumatoid arthritis,
CC   HIV infection and other immunological and inflammatory diseases.
SQ   Sequence   50 AA;


  Query Match             53.4%;  Score 39.5;  DB 1;  Length 50;
  Best Local Similarity   43.8%;  Pred. No. 11;
  Matches    7;  Conservative    4;  Mismatches    4;  Indels    1;  Gaps    1;

Qy       1 QHCDSSECKSSP-RCK 15
           | |  ::|| :| ||:
Db      13 QLCVDGQCKCTPIRCR 28


RESULT   4
R95596
ID   R95596 standard; protein; 139 AA.
AC   R95596;
DT   17-DEC-1996  (first entry)
DE   Factor X light chain.
KW   Factor X; light chain; human; heavy chain; Factor Xa; prothrombinase;
KW   activating enzyme; blood factor; immunoaffinity chromatography; therapy;
KW   antigen; inhibitor; activated resin; coagulation disorder; factor II;
KW   vasculature function disorder; factor V; factor IX; factor XI; protein C;
KW   factor XII; factor VII; protein S; fibrinogen.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   disulfide_bond  17. .22
FT   domain          17. .22
FT                   /note= "GLA domain"
FT   disulfide_bond  50. .61
FT   domain          50. .124
FT                   /note= "growth factor domains"
FT   disulfide_bond  55. .70
FT   misc_difference 63
FT                   /note= "unspecified amino acid, represented in
FT                   specification as beta"
FT   disulfide_bond  72. .81
FT   disulfide_bond  89. .100
FT   disulfide_bond  96. .109
FT   disulfide_bond  111. .124
FT   disulfide_bond  132
FT                   /note= "disulphide bond to residue 160 of Factor X heavy
FT                   chain (see W05820), or to residue 108 of Factor
FT                   Xa heavy chain (see R95597 and R95598)"
PN   WO9613274-A1.
PD   09-MAY-1996.
PF   27-OCT-1995; U13940.
PR   28-OCT-1994; US-330978.
PA   (CORT-) COR THERAPEUTICS INC.
PI   King RS;
DR   WPI; 96-239270/24.
PT   Prepn. of an inhibited form of an activated blood factor, e.g.
PT   factor X - by treating partially purified blood factor prepns. with
PT   an activating factor and an inhibiting factor
PS   Disclosure; Fig 1; 45pp; English.
CC   This sequence represents the light chain of human Factor X (the heavy
CC   chain is represented by W05820).  Factor X must be activated to Factor Xa
CC   before the protease is incorporated into the prothrombinase complex.  In
CC   Factor Xa the light chain sequence is identical to the Factor X light
CC   chain, and the heavy chain is a truncated version of the Factor X heavy
CC   chain.  An inhibited form of activated Factor X is prepared by the method
CC   of the invention.  In this method, a partially purified preparation
CC   containing the blood factor is treated to convert the factor into an
CC   activated form (using an immobilised activating enzyme).  The activated
CC   form is then converted into an inhibited form in a single step, and the
CC   inhibited factor is recovered.  The inhibited blood factor is recovered
CC   by immunoaffinity chromatography using an antigen specific monoclonal
CC   antibody coupled to an activated resin (such as agarose), or an anion
CC   exchange column with an anion-exchange group linked to a naturally
CC   derived polysaccharide or a synthetically derived polymeric matrix.  The
CC   activated resin used preferably uses activation chemistry selected from
CC   tresyl, azactone, aldehyde, hydrazide, N-hydroxy succinimide or triazine.
CC   This method produces a highly purified preparation of an inhibited form
CC   (either permanently or transiently inhibited) of an activated blood
CC   factor in high yield.  The factors produced can be used in the treatment
CC   of coagulation disorders, or disorders of vasculature function.
SQ   Sequence   139 AA;


  Query Match             56.8%;  Score 42;  DB 1;  Length 139;
  Best Local Similarity   33.3%;  Pred. No. 12;
  Matches    5;  Conservative    6;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 QHCDSSECKSSPRCK 15
           ::|::| | :  :||
Db      48 DQCETSPCQNQGKCK 62
```

```
RESULT   5
W40283
ID   W40283 standard; Protein; 467 AA.
AC   W40283;
DT   16-JUN-1998  (first entry)
DE   Human Factor X protease.
KW   Factor X; factor IX; serine protease activity; catalytic domain; ZAD;
KW   zymogen-activating domain; epidermal growth factor-like domain; EGF1;
KW   EGF2; regulator; coagulation; fibrinolysis; homeostasis; X-ray structure;
KW   detection; drug modelling; restriction protease.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Domain          108. .153
FT                   /label= EGF2 domain
FT   Domain          154. .165
FT                   /label= EGF2 domain
FT   Domain          166. .216
FT                   /label= Activating domain
FT   Domain          217. .454
FT                   /label= catalytic domain
PN   WO9747737-A1.
PD   18-DEC-1997.
PF   11-JUN-1997; E03027.
PR   06-JUL-1996; EP-110959.
PR   11-JUN-1996; EP-109288.
PR   22-JUN-1996; EP-110109.
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
PI   Hopfner K, Kopetzki E;
DR   WPI; 98-052304/05.
DR   N-PSDB; V10462.
PT   Non-glycosylated, truncated forms of factor IX family protein with
PT   serine protease activity - used to screen for specific modulators
PT   and to assay factor IXa
PS   Disclosure; Fig 3; 49pp; German.
CC   This sequence represents a human factor X protease. This protein is used
CC   in the construction of a novel non-glycosylated protein and truncated
CC   and zymogen forms of this protein, which have serine protease activity.
CC   The protein is composed of various domains from a factor IX family
CC   protein, namely a catalytic domain (CD) N-terminally bound to a
CC   zymogen-activating domain (ZAD), N-terminally bound to an EGF1 and/or
CC   EGF2 domain (EGF = epidermal growth factor-like domain). Such proteins
CC   are used to identify activators/inhibitors of factor IX family proteins
CC   (potentially useful as regulators of coagulation, fibrinolysis and
CC   homeostasis). The protein in zymogen form is also useful in assays for
CC   detecting factor IXa activity in aqueous solution (specifically in body
CC   fluids). The protein can be used to produce co-crystals with protease
CC   variants or inhibitors for X-ray structural analysis and drug modelling
CC   and as restriction proteases in biotechnology. These truncated proteins
CC   have the same specificity as factor IX family proteases and can be
CC   produced in prokaryotes in a form that allows production of active enzyme
CC   by conversion to native form and enzymatic cleavage.
SQ   Sequence   467 AA;

  Query Match             59.5%;  Score 44;  DB 1;  Length 467;
  Best Local Similarity   33.3%;  Pred. No. 17;
  Matches    5;  Conservative    7;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 QHCDSSECKSSPRCK 15
          ::|::| | : ::||
Db     70 DQCETSPCQNQAKCK 84


RESULT   6
R13675
ID   R13675 standard; Protein; 250 AA.
AC   R13675;
DT   09-OCT-1991  (first entry)
DE   Factor X-LACI hybrid protein.
KW   Kunitz domain; blood; coagulation; inhibitor; Factor X;
KW   tissue factor; TF; Lipoprotein-Associated Coagulation Inhibitor.
FH   Key             Location/Qualifiers
FT   protein         1. .171
FT                   /label= Xlc
FT   peptide         1. .40
FT                   /label= prepro_leader
FT   domain          55. .64
FT                   /label= GLA_domain
FT   domain          89. .150
FT                   /label= growth_factor_domains
FT   domain          172. .250
FT                   /label= kunitz_domain
FT   disulfide_bond  57. .62
FT   disulfide_bond  90. .101
FT   disulfide_bond  95. .110
FT   disulfide_bond  112. .121
FT   disulfide_bond  129. .140
FT   disulfide_bond  136. .149
FT   disulfide_bond  151. .164
FT   disulfide_bond  186. .236
FT   disulfide_bond  195. .219
FT   disulfide_bond  211. .232
PN   EP-439442-A.
PD   31-JUL-1991.
PF   21-JAN-1991; 870008.
PR   25-JAN-1990; US-470289.
PA   (UNIW ) UNIV OF WASHINGTON.
PI   Girard TJ, Broze GJ;
DR   WPI; 91-224839/31.
DR   N-PSDB; Q12776.
PT   New factor X-LACI hybrid protein - comprises light chain of
PT   factor X and LACI's first kunitz domain for use as anticoagulant
PS   Disclosure; Page 12-14; 17pp; English.
CC   The protein is used as a blood coagulation inhibitor in mammals.  It
CC   is believed to mimic the Xa/LACI complex in binding to and
CC   inhibiting VIIa/tissue factor.  LACI inhibits via a novel feedback
CC   mechanism requiring generation of Xa (a prod. of VIIa/TF activity);
CC   XlcLACIki inhibits VIIa/TF activity directly.
CC   The DNA allows prodn. of XlcLACIki by introduction of the gene into
CC   cells suitable for expression, e.g. E. coli or CHO cells.
SQ   Sequence   250 AA;

  Query Match             56.8%;  Score 42;  DB 1;  Length 250;
  Best Local Similarity   33.3%;  Pred. No. 19;
  Matches    5;  Conservative    6;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 QHCDSSECKSSPRCK 15
          ::|::| | :  :||
Db     88 DQCETSPCQNQGKCK 102


RESULT   7
R32501
ID   R32501 standard; protein; 400 AA.
AC   R32501;
DT   09-JUN-1993  (first entry)
DE   Beta-adrenergic receptor.
KW   Fat cell specific; BAR; lipolysis; obesity; diagnosis;
KW   thermogenesis; metabolism.
OS   Rattus rattus.
PN   US7783602-A.
PD   15-DEC-1992.
PF   11-NOV-1991; 783602.
PR   01-NOV-1991; US-783602.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI   Venter CJ;
DR   WPI; 93-067426/08.
PT   Fat cell specific beta- adrenergic receptor polypeptide - used
PT   for diagnosis of obesity due to inactive lipolysis
PS   Disclosure; Page 16; 20pp; English.
CC   A rat intercapsular brown adipose tissue cDNA library was cloned and
CC   probes with DNA probes encoding human beta-1 and rat beta-2
CC   adrenergic receptors under low stringency conditions.  Positive
CC   clones were found to be different from the rat and human sequences
CC   and contained a single open reading frame of 1200 bp encoding the
```

```
CC   protein shown, of 400 amino acids and mol. wt. 43 kD.  The protein is
CC   the fat specific beta-adrenergic receptor and may be used in work on
CC   the thermogenesis process.  Isolation of the gene for BAR allows the
CC   diagnosis and treatment of obesity and the testing of medications
CC   for their effectiveness in stimulating the thermogenesis metabolic
CC   response in obesity patients.
SQ   Sequence   400 AA;


  Query Match             58.1%;  Score 43;  DB 1;  Length 400;
  Best Local Similarity   58.3%;  Pred. No. 21;
  Matches    7;  Conservative    3;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 DSSECKSSPRCK 15
           :: || |:||||
Db     182 EAQECHSNPRCK 193


RESULT   8
W80576
ID   W80576 standard; Protein; 52 AA.
AC   W80576;
DT   16-DEC-1998  (first entry)
DE   Human factor VII EGF domain primary sequence.
KW   O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW   O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
OS   Synthetic.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Peptide         44. .49
FT                   /note= "histidine tag"
PN   WO9833924-A1.
PD   06-AUG-1998.
PF   17-DEC-1997; U23401.
PR   26-NOV-1997; US-978741.
PR   31-JAN-1997; US-792498.
PA   (GETH ) GENENTECH INC.
PI   Spellman MW, Wang Y;
DR   WPI; 98-437477/37.
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth
PT   factor domains - useful for diagnosis and treatment of diseases
PT   involving overexpression of the enzyme
PS   Example 2; Page 39; 90pp; English.
CC   This represents the primary sequence of the first EGF domain from human
CC   factor VII contained in a plasmid construct. The invention provides a
CC   human heart O-fucosyltransferase enzyme that can glycosylate an epidermal
CC   growth factor (EGF) domain of a polypeptide with an activated O-fucose
CC   residue. Inhibitors of O-fucosyltransferase, e.g. mutants with increased
CC   affinity for the EFG domains, are used in diagnosis and treatment of
CC   conditions associated with overexpression of O-fucosyltransferase, to
CC   promote survival of sensory (retinal) neurons. Probes based on EGF domain
CC   polypeptide are used to detect gene amplification and expression. The
CC   expression can also be determined at the protein level using antibodies
CC   specific for O-fucosyltransferase.
SQ   Sequence   52 AA;


  Query Match             50.0%;  Score 37;  DB 1;  Length 52;
  Best Local Similarity   33.3%;  Pred. No. 26;
  Matches    5;  Conservative    6;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 QHCDSSECKSSPRCK 15
           ::|:|: | ::  ||
Db       5 DQCESNPCLNGGSCK 19


RESULT   9
R40210
ID   R40210 standard; protein; 53 AA.
AC   R40210;
DT   04-FEB-1994  (first entry)
DE   Sequence of yeast class II metallothionine.
KW   Metallothionine; yeast; class II.
OS   Saccharomyces cervisiae.
PN   DE4212134-A.
PD   19-AUG-1993.
PF   10-APR-1992; 212134.
PR   17-FEB-1992; GB-003299.
PA   (INDE-) INDENA SPA.
PI   Bombardelli E,  Ponzone C,  Puglisi PP;
DR   WPI; 93-265710/34.
PT   Topical compsn. for protecting tissue e.g. skin - against toxic
PT   heavy metals, contg. metal-complexing protein with high cysteine
PT   content
PS   Disclosure; Page 3; 7pp; German.
CC   Class II metallothionines have the same potential to form
CC   thio-metal  complexes and clusters as Class II metallotionines.
CC   This class comprises metallothionines whose primary structure
CC   bears only a slight, or no, resemblance to those of mammals.
SQ   Sequence   53 AA;


  Query Match             50.0%;  Score 37;  DB 1;  Length 53;
  Best Local Similarity   30.8%;  Pred. No. 27;
  Matches    4;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 HCDSSECKSSPRC 14
           :|: : ||:: :|
Db       8 QCQCGSCKNNEQC 20


RESULT  10
R37437
ID   R37437 standard; Protein; 39 AA.
AC   R37437;
DT   30-SEP-1993  (first entry)
DE   Thrombin inhibitor C-terminal (AS 19-57).
KW   Primer; PCR; thrombosis; arterial reocclusion; blood; thrombin;
KW   hirudin.
OS   Haemadipsa sylvestris.
PN   DE4209110-A.
PD   27-MAY-1993.
PF   20-MAR-1992; 209110.
PR   26-NOV-1991; DE-138698.
PR   20-MAR-1992; DE-209110.
PA   (BADI ) BASF AG.
PI   Bialojan S,  Friedrich T,  Kroeger B,  Strube K;
DR   WPI; 93-176720/22.
DR   N-PSDB; Q42433.
PT   New proteins obtd. from Haemadipsa Sylvestris - used as thrombin
PT   inhibitors for treatment and prevention of thrombosis and
PT   arterial re-occlusion
PS   Example 6; Page 8; 14pp; German.
CC   Example 6 describes the cloning of DNA encoding the thrombin
CC   inhibiting protein.  A cDNA library is produced from H. sylvestris
CC   DNA.  Primers (i - Q42426) and (ii - Q42427) were based on peptide
CC   fragments corresp. to amino acids 4-11 and 19-28 respectively.
CC   The 3' primers were A-B-T18 (Q42428), A (derived from A-B-T18 -
CC   Q42429) and B (derived from A-B-T18 - Q42430).  A first sequence
CC   was obtained (Q42433).  A further PCR cycle was performed using
CC   primers (iii - Q42431) and (iv - Q42432).  The sequence
CC   given in Q42434 was obtained.
SQ   Sequence   39 AA;


  Query Match             48.6%;  Score 36;  DB 1;  Length 39;
  Best Local Similarity   30.8%;  Pred. No. 29;
  Matches    4;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 QHCDSSECKSSPR 13
           | |: ::| : |:
Db      12 QSCNDGQCSGDPK 24


RESULT  11
P94625
```

```
ID   P94625 standard; protein; 61 AA.
AC   P94625;
DT   22-JUN-1990  (first entry)
DE   CUP1 translation product of plasmid M13mp18/CUP1.
KW   Copper resistance gene; expression system; yeast; cassette; ds.
OS   Saccharomyces cerevisiae
PN   AU8815845-A.
PD   10-NOV-1988.
PF   7-MAY-1987; 15845.
PR   7-MAY-1987; AU-015845.
PR   6-MAY-1988; AU-001788.
PA   (CSIR) Commonwealth Scient Org.
PI   Macreadie IG, Vaughan PR, Azad AA;
DR   WPI; 89-069074/10.
DR   N-PSDB; N90634.
PT   Gene expression system -
PT   contains yeast copper resistance gene and multiple cloning site
PT   and is movable as a cassette.
PS   Disclosure; p; English.
CC   Expression vector carrying the yeast copper resistance genes allows for
CC   the expression of foreign proteins with a copper induced promoter and in
CC   the absence of their own start codon.
SQ   Sequence   61 AA;


  Query Match             50.0%;  Score 37;  DB 1;  Length 61;
  Best Local Similarity   30.8%;  Pred. No. 30;
  Matches    4;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 HCDSSECKSSPRC 14
           :|: : ||:: :|
Db      16 QCQCGSCKNNEQC 28


RESULT  12
P50776
ID   P50776 standard; Protein; 61 AA.
AC   P50776;
DT   30-NOV-1991  (first entry)
DE   Sequence encoded by yeast copper metallothionein gene in
DE   YEp36 insert.
KW   Metal poisoning therapy; chelating agent;
KW   industrial waste treatment.
OS   Saccharomyces cerevisiae.
PN   EP-134773-A.
PD   20-MAR-1985.
PF   07-AUG-1984; 870112.
PR   08-AUG-1983; US-520668.
PR   02-MAR-1984; US-584657.
PA   (SMIK ) SMITHKLINE BECKMAN CORP.
PI   Butt TR;
DR   WPI; 85-070084/12.
DR   N-PSDB; N50529.
PT   Gene coding for yeast copper metallo:thionein - isolated from
PT   yeast by hybridisation and useful for micro-organism
PT   transformation
PS   Disclosure; Page 4-6; 25pp; English.
CC   The gene is present in a Sau 3A-Sau 3A fragment of chromosomal DNA
CC   in a Cu-resistant (1mM CuSO4) mutant. The promoter region is present
CC   in a Sau 3A-Kpn I, Xbal-KpnI or Rsa I-Kpn I fragment. Recombinant
CC   DNA molecules contg. the structural gene include YEp1, YEp2, YEp29,
CC   YEp36 and TRp20, (all claimed).
SQ   Sequence   61 AA;


  Query Match             50.0%;  Score 37;  DB 1;  Length 61;
  Best Local Similarity   30.8%;  Pred. No. 30;
  Matches    4;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 HCDSSECKSSPRC 14
           :|: : ||:: :|
Db      16 QCQCGSCKNNEQC 28


RESULT  13
P30006
ID   P30006 standard; peptide; 61 AA.
AC   P30006;
DT   25-APR-1992  (first entry)
DE   Sequence of yeast copper chelating metallothionein
DE   Cu-chelatin).
KW   Copper chelating metallothionein; heavy metal; copper-resistant;
KW   yeast vector; industrial waste.
OS   Saccharomyces cerevisiae strain BZ31-1-7Ba.
PN   EP--96491-A.
PD   21-DEC-1983.
PF   03-JUN-1983; 098143.
PR   03-JUN-1982; US-384821.
PA   (REGC ) UNIV OF CALIFORNIA.
PI   Fogel S, Welch JW, Karin M;
DR   WPI; 83-846233/51.
DR   N-PSDB; N30011.
PT   Pure yeast copper chelatin - prodn. by recombinant DNA methods
PT   from DNA sequences
PS   Disclosure; Page 7; 16pp; English.
CC   The inventors claim yeast copper chelatin, its fragments and
CC   analogues and DNA encoding them.  Cu-chelatin is useful as a
CC   chelating agent for removal of heavy metals, esp. copper, from
CC   industrial wastes, etc.  Plasmids isolated from hosts transformed to
CC   copper resistant are useful in prepn. of hybridisation probes.  A
CC   yeast expression system contg. the Cu-chelatin DNA and upstream
CC   regulatory sequence together with a foreign gene and regulatory
CC   signals is also claimed.
SQ   Sequence   61 AA;


  Query Match             50.0%;  Score 37;  DB 1;  Length 61;
  Best Local Similarity   30.8%;  Pred. No. 30;
  Matches    4;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 HCDSSECKSSPRC 14
           :|: : ||:: :|
Db      16 QCQCGSCKNNEQC 28


RESULT  14
R54087
ID   R54087 standard; Protein; 61 AA.
AC   R54087;
DT   01-FEB-1995  (first entry)
DE   Metallothionein.
KW   Metallothionein; CUP1; promoter; yeast; copper; expression;
KW   hirudin.
OS   Synthetic.
PN   EP-603128-A.
PD   22-JUN-1994.
PF   07-DEC-1993; 810853.
PR   15-DEC-1992; EP-811005.
PA   (CIBA ) CIBA GEIGY AG.
PA   (UCPG-) UCP GEN-PHARMA AG.
PI   Fuerst P,  Heim J,  Hottiger T,  Pohlig G;
DR   WPI; 94-193627/24.
DR   N-PSDB; Q64145.
PT   Production of polypeptide(s) with improved stability and
PT   increased yields - using engineered yeast cells transformed with
PT   plasmids carrying a CUP1 gene, and adding copper to the culture
PT   medium.
PS   Disclosure; Page 20-21; 36pp; English.
CC   Example 1 describes the construction of plasmid pHE112R, a 2 micron
CC   plasmid contg. the GAPFLp-hirudin expression cassette and the full
CC   CUP1 gene.  A 1.3 kb BamHI fragment contg. the full metallothionein
CC   encoding gene - CUP1, is isolated from plasmid YEp3362xSst (Wright,
CC   C.F et al., Nucleic Acids Res. 14 (1986), 8489-8499).  YEp3362xSst
CC   is digested with BamHI, the 1.3 kb fragment isolated, purified and
CC   ligated with BamHI cut pUC19.  The resulting plasmid is named pHE105.
CC   The 1.3 kb fragment (see Q64145) is ligated into SnaBI-cut
```

```
CC   pDP34/GAPFL-YHIR.  E. coli is transformed with the resulting
CC   plasmid pHE112R.
SQ   Sequence   61 AA;


  Query Match             50.0%;  Score 37;  DB 1;  Length 61;
  Best Local Similarity   30.8%;  Pred. No. 30;
  Matches    4;  Conservative    6;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 HCDSSECKSSPRC 14
          :|: : ||:: :|
Db     16 QCQCGSCKNNEQC 28


RESULT  15
R22513
ID   R22513 standard; Protein; 436 AA.
AC   R22513;
DT   28-JUL-1992  (first entry)
DE   Truncated precursor of human Factor Xai.
KW   Mutant; prothrombinase complex; proteolylic; precursor; thrombosis;
KW   inflammation; restenosis; transplantation; haemophilia; antibodies.
OS   Homo sapiens.
PN   WO9204378-A.
PD   19-MAR-1992.
PF   04-SEP-1991; U06337.
PR   04-SEP-1990; US-578646.
PA   (CORT-) COR THERAPEUTICS IN.
PI   Wolf D;
DR   WPI; 92-114303/14.
PT   New analogues of Factor XA peptide - useful for treating
PT   haemophilia, thrombosis, inflammation and transplant
PT   complications, for in-vivo diagnosis
PS   Claim 7; Fig 1; 59pp; English.
CC   The full length cDNA of human factor X (Mpl9X) was converted to
CC   encode a truncated form of human Factor X, designated rX', by deletion
CC   of the activation peptide by oligonucleotide site directed mutagenesis.
CC   An oligonucleotide was used to align Arg 142 following the
CC   C-terminus of the Factor X light chain with Ile 53 of the Factor X
CC   activation peptide (1st residue of the heavy chain). When expressed
CC   in CHO cells the truncated peptide was cleaved endogenously.
CC   Modified Factor Xa was further produced by acylation e.g. with the
CC   p-nitrophenyl ester of p-toluoylic acid.  Factor rX' is used to
CC   treat or prevent thrombosis; inflammation; restenosis or complications
CC   of transplantation.  It is also used in treatment of adult respiratory
CC   distress syndrome and haemophilia.  The modified Factor X has no
CC   proteolytic activity and interferes with the ability of endogenous
CC   factor Xa to convert prothrombin to thrombin.  Antibodies reactive
CC   with Factor rX' are passive therapeutic agents and used for diagnosis.
CC   See also R22512.
SQ   Sequence   436 AA;


  Query Match             56.8%;  Score 42;  DB 1;  Length 436;
  Best Local Similarity   33.3%;  Pred. No. 31;
  Matches    5;  Conservative    6;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 QHCDSSECKSSPRCK 15
          ::|::| | :  :||
Db     88 DQCETSPCQNQGKCK 102


Search completed: May  4, 1999, 12:32:39
Job time: 9786 sec
```

OM protein - protein search, using sw model

Run on: May 4, 1999, 06:22:08 ; Search time 25.07 Seconds
(without alignments)
243.559 Million cell updates/sec

Title: US-09-037-460-2_COPY_1_163
Perfect score: 820
Sequence: 1 MKSVLLLTTLLVPAHLVAAW.....NRFVSLTEHDMASGDGNIVR 163

Scoring table: PAM150

Searched: 116738 seqs, 37460341 residues

Database : PIR_58:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	141.5	17.3	351	2	S20078	NOV protein - chic
2	123.5	15.1	357	2	I38069	gene novH protein
3	115.5	14.1	258	2	A45403	insulin-like growt
4	114.5	14.0	271	2	I48604	insulin-like growt
5	112	13.7	348	2	A40578	beta IG-M2 protein
6	112	13.7	348	2	A53228	fisp-12 protein pr
7	110.5	13.5	271	2	JC1463	insulin-like growt
8	111	13.5	375	2	A41428	CEF-10 protein pre
9	109.5	13.4	272	2	A53748	insulin-like growt
10	109	13.3	349	2	A40551	connective tissue
11	107	13.0	237	2	I47031	insulin-like growt
12	106.5	13.0	258	2	G01662	insulin-like growt
13	106.5	13.0	258	2	B37252	insulin-like growt
14	104.5	12.7	266	2	A35037	insulin-like growt
15	104.5	12.7	271	2	JC4584	insulin-like growt
16	103.5	12.6	254	2	I48599	insulin-like growt
17	103.5	12.6	254	2	JC1464	insulin-like growt
18	103.5	12.6	379	2	A35669	gene CYR61 protein
19	101.5	12.4	291	1	IOHU3	insulin-like growt
20	101.5	12.4	291	1	JN0064	insulin-like growt
21	100.5	12.3	291	2	I48602	insulin-like growt
22	98.5	12.0	292	2	A36748	insulin-like growt
23	98	12.0	251	2	A55035	cysteine-rich prot
24	103	12.6	1801	1	MMRTS	laminin beta-2 cha
25	101.5	12.4	1138	1	S24066	protein-tyrosine k
26	97	11.8	310	2	A60967	insulin-like growt
27	97	11.8	317	2	I46916	insulin-like growt
28	94.5	11.5	259	1	IOHU1	insulin-like growt
29	90.5	11.0	81	2	A45320	transglutaminase s
30	97.5	11.9	1136	1	S57845	protein-tyrosine k
31	94.5	11.5	1134	1	JN0711	protein-tyrosine k
32	87.5	10.7	111	2	B45403	insulin-like growt
33	94.5	11.5	1786	1	MMMSB1	laminin beta-1 cha
34	93.5	11.4	1816	1	S68960	laminin alpha-4 ch
35	92.5	11.3	1797	2	A55677	laminin beta-2 cha
36	92	11.2	1786	1	MMHUB1	laminin beta-1 cha
37	86.5	10.5	328	1	A41927	insulin-like growt
38	85.5	10.4	254	2	I48603	insulin-like growt
39	94	11.5	5147	1	IJFFTM	cadherin-related t
40	88	10.7	728	2	I50719	C-Delta-1 - chicke
41	90	11.0	1639	1	MMFFB2	laminin gamma-1 ch
42	84.5	10.3	282	2	S50031	prostacyclin-stimu
43	84.5	10.3	304	2	A33274	insulin-like growt
44	84.5	10.3	305	2	I48601	insulin-like growt
45	84.5	10.3	305	2	JN0508	insulin-like growt

ALIGNMENTS

RESULT 1
S20078
NOV protein - chicken
C;Species: Gallus gallus (chicken)
C;Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 10-Sep-1997
C;Accession: S20078
R;Joliot, V.; Martinerie, C.; Dambrine, G.; Plassiart, G.; Brisac, M.; Crochet, J.; P
Mol. Cell. Biol. 12, 10-21, 1992
A;Title: Proviral rearrangements and overexpression of a new cellular gene (nov) in m
A;Reference number: S20078; MUID:92107157
A;Accession: S20078
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-351 <JOL>
A;Cross-references: EMBL:X59284; NID:g63702; PID:g63703
C;Genetics:
A;Gene: NOV

Query Match 17.3%; Score 141.5; DB 2; Length 351;
Best Local Similarity 31.9%; Pred. No. 3.1e-08;
Matches 45; Conservative 17; Mismatches 38; Indels 41; Gaps 7;

```
Qy       3 SVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAA 59
           :||||  ||: :  |:::          ||: | :     :|||     :||| |||| |||
Db      10 PVLLLLLLLLRPCEVSGR----EAACPRPCGGRCPAEPPRCAPGVPAVLDGCGCCLVCAR 65

Qy      60 GRGETCYRTVSGMDGMKCGPGLRCQPSNG-------EDPFGEEFGICK-----DCPYG-- 106
            |||:              |:| | |: | |       ||  |   |||      :| :
Db      66 QRGES-----------CSPLLPCDESGGLYCDRGPED--GGGAGICMVLEGDNCVFDGM 111

Qy     106 ------TFGMDCRETCNCQSG 120
                 ||   |:  | |: |
Db     112 IYRNGETFQPSCKYQCTCRDG 132
```

RESULT 2
I38069
gene novH protein - human
C;Species: Homo sapiens (man)
C;Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 28-Feb-1997
C;Accession: I38069
R;Martinerie, C.; Huff, V.; Joubert, I.; Badzioch, M.; Saunders, G.; Strong, L.; Perb
Oncogene 9, 2729-2732, 1994
A;Title: Structural analysis of the human nov proto-oncogene and expression in Wilms
A;Reference number: I38069; MUID:94336229
A;Accession: I38069
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-357 <RES>
A;Cross-references: EMBL:X78351; NID:g587422; PID:g825696
C;Genetics:
A;Gene: novH
A;Introns: 28/3; 104/1; 188/1; 259/3

Query Match 15.1%; Score 123.5; DB 2; Length 357;
Best Local Similarity 31.8%; Pred. No. 2.8e-06;
Matches 42; Conservative 25; Mismatches 34; Indels 31; Gaps 9;

```
Qy       7 LTTLLVPAHLVA--AWSNNYAVDCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGR 61
           || ||:  ||::  | :   : :|| :| :    ::| |    |:||| |:|| |||  |
```

```
Db      18 LTFLLL--HLLGQVAATQRCPPQCPGRCPA----TPPTCAPGVRAVLDGCSCCLVCARQR 71

Qy      62 GETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICK-----DCPYG--------TFG 108
           ||:|    | ::    ::|| |: |   || ::: |||      :| :          |
Db      72 GESC----SDLEPCDESSGLYCDRS--ADP-SNQTGICTAVEGDNCVFDGVIYRSGEKFQ 124

Qy     109 MDCRETCNCQSG 120
             |:  | |: |
Db     125 PSCKFQCTCRDG 136


RESULT   3
A45403
insulin-like growth factor-binding protein 4 - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 01-Dec-1995
C;Accession: A45403
R;Moser, D.R.; Lowe Jr., W.L.; Dake, B.L.; Booth, B.A.; Boes, M.; Clemmons, D.R.; Bar, R
Mol. Endocrinol. 6, 1805-1814, 1992
A;Title: Endothelial cells express insulin-like growth factor-binding proteins 2 to 6.
A;Reference number: A45403; MUID:93125553
A;Accession: A45403
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-258 <MOS>
A;Experimental source: pulmonary artery endothelial cells
A;Note: sequence extracted from NCBI backbone (NCBIN:122183, NCBIP:122184)
C;Superfamily: thyroglobulin type I repeat homology
F;174-249/Domain: thyroglobulin type I repeat homology <THY1>


  Query Match             14.1%;  Score 115.5;  DB 2;  Length 258;
  Best Local Similarity   32.3%;  Pred. No. 1.6e-05;
  Matches   30;  Conservative   24;  Mismatches   30;  Indels    9;  Gaps    4;

Qy       1 MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDS--SECKSSPRCKRTVLD-DCGCCRVC 57
           | |: |:::||::|   :: ::: |: ||   :   : |:::  |   | :  ||||  |
Db     * 1 MLSLCLMAALLLAAGPGPSLGDE-AIHCPPCSEEKLARCRPPVGCEELVREPGCGCCATC 59

Qy      58 AAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           | |:|  |     |:   :||:|||| |: | :
Db      60 ALGKGMPC-----GVYTPRCGSGLRCYPPRGVE 87


RESULT   4
I48604
insulin-like growth factor binding protein 5 precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 10-Sep-1997
C;Accession: I48604; A48699; A54259
R;Schuller, A.G.; Groffen, C.; van Neck, J.W.; Zwarthoff, E.C.; Drop, S.L.
Mol. Cell. Endocrinol. 104, 57-66, 1994
A;Title: cDNA cloning and mRNA expression of the six mouse insulin-like growth factor bi
A;Reference number: I48600; MUID:95121750
A;Accession: I48604
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-271 <RES>
A;Cross-references: EMBL:X81583; NID:g550384; PID:g550385
R;James, P.L.; Jones, S.B.; Busby Jr., W.H.; Clemmons, D.R.; Rotwein, P.
J. Biol. Chem. 268, 22305-22312, 1993
A;Title: A highly conserved insulin-like growth factor-binding protein (IGFBP-5) is expr
A;Reference number: A48699
A;Accession: A48699
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-271 <JAM>
A;Cross-references: GB:L12447; NID:g425405; PID:g293384
R;Kou, K.; Jenkins, N.A.; Gilbert, D.J.; Copeland, N.G.; Rotwein, P.
Genomics 20, 412-418, 1994
A;Title: Organization, expression, and chromosomal location of the mouse insulin-like gr
A;Reference number: A54259
A;Accession: A54259
A;Molecule type: DNA
A;Residues: 1-111 <KOU>
A;Cross-references: GB:U02023
C;Genetics:
A;Gene: IGFBP-5
C;Superfamily: thyroglobulin type I repeat homology
F;191-262/Domain: thyroglobulin type I repeat homology <THY1>


  Query Match             14.0%;  Score 114.5;  DB 2;  Length 271;
  Best Local Similarity   28.7%;  Pred. No. 2.2e-05;
  Matches   27;  Conservative   20;  Mismatches   26;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::  |||: :::           ||:: : |: : |  : |         ||||
Db       7 LLLLAAYAVPAQGLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 57

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      58 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 86


RESULT   5
A40578
beta IG-M2 protein precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 06-Mar-1992 #sequence_revision 06-Mar-1992 #text_change 20-Mar-1998
C;Accession: A40578
R;Brunner, A.; Chinn, J.; Neubauer, M.; Purchio, A.F.
DNA Cell Biol. 10, 293-300, 1991
A;Title: Identification of a gene family regulated by transforming growth factor-beta
A;Reference number: A40578; MUID:91229699
A;Accession: A40578
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-348 <BRU>
A;Cross-references: GB:M80263; NID:g201945; PID:g201946


  Query Match             13.7%;  Score 112;  DB 2;  Length 348;
  Best Local Similarity   30.4%;  Pred. No. 4.8e-05;
  Matches   34;  Conservative   15;  Mismatches   35;  Indels   28;  Gaps    7;

Qy      27 DCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGP--GL 81
           ||: :|: :   ::|:|       ||| |||||||   || |        :   | |  ||
Db      27 DCSAQCQCAA-EAAPHCPAGVSLVLDGCGCCRVCAKQLGELC------TERDPCDPHKGL 79

Qy      82 RCQPSNGEDPFGEEFGICKD-----CPYG--------TFGMDCRETCNCQSG 120
             |: ::   |     |:|       | :|         :|   |:  | |  |
Db      80 FCDFGS---PANRKIGVCTAKDGAPCVFGGSVYRSGESFQSSCKYQCTCLDG 128


RESULT   6
A53228
fisp-12 protein precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-May-1994 #sequence_revision 19-May-1994 #text_change 10-Sep-1997
C;Accession: A53228
R;Ryseck, R.P.; Macdonald-Bravo, H.; Mattei, M.G.; Bravo, R.
Cell Growth Differ. 2, 225-233, 1991
A;Title: Structure, mapping, and expression of fisp-12, a growth factor-inducible gen
A;Reference number: A53228
A;Accession: A53228
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-348 <RYS>
A;Cross-references: GB:M70641; NID:g193313; PID:g193314
C;Genetics:
A;Gene: fisp-12


  Query Match             13.7%;  Score 112;  DB 2;  Length 348;
  Best Local Similarity   30.4%;  Pred. No. 4.8e-05;
```

```
  Matches   34;  Conservative   15;  Mismatches   35;  Indels   28;  Gaps    7;

Qy      27 DCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGP--GL 81
           ||: :|: :   ::|:|     ||| ||||||||   || |       :   | |  ||
Db      27 DCSAQCQCAA-EAAPHCPAGVSLVLDGCGCCRVCAKQLGELC------TERDPCDPHKGL 79

Qy      82 RCQPSNGEDPFGEEFGICKD-----CPYG--------TFGMDCRETCNCQSG 120
             |: ::   |     |:|       | :|        :|   |:  | |  |
Db      80 FCDFGS---PANRKIGVCTAKDGAPCVFGGSVYRSGESFQSSCKYQCTCLDG 128


RESULT   7
JC1463
insulin-like growth factor-binding protein 5 precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 10-Sep-1997
C;Accession: JC1463; A40403; F40403
R;Zhu, X.; Ling, N.; Shimasaki, S.
Biochem. Biophys. Res. Commun. 190, 1045-1052, 1993
A;Title: Cloning of the rat insulin- like growth factor binding protein-5 gene and DNA s
A;Reference number: JC1463
A;Accession: JC1463
A;Molecule type: DNA
A;Residues: 1-271 <ZHU>
A;Cross-references: GB:L08275
R;Shimasaki, S.; Shimonaka, M.; Zhang, H.P.; Ling, N.
J. Biol. Chem. 266, 10646-10653, 1991
A;Title: Identification of five different insulin-like growth factor binding proteins (I
A;Reference number: A40403; MUID:91244847
A;Accession: A40403
A;Molecule type: mRNA
A;Residues: 1-271 <SHI>
A;Cross-references: GB:M62781; NID:g204745; PID:g204746
A;Accession: F40403
A;Status: preliminary
A;Molecule type: protein
A;Residues: 20-25,'X',27-28,'X',30-36,'X',38-43,'X',45-51,'X',53,'XX' <SH2>
C;Genetics:
A;Introns: 112/3; 188/3; 228/3
C;Superfamily: thyroglobulin type I repeat homology
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-271/Product: insulin-like growth factor binding protein 5 #status predicted <MAT>
F;191-262/Domain: thyroglobulin type I repeat homology <THY1>

  Query Match             13.5%;  Score 110.5;  DB 2;  Length 271;
  Best Local Similarity   28.7%;  Pred. No. 5.8e-05;
  Matches   27;  Conservative   20;  Mismatches   26;  Indels   21;  Gaps    3;

Qy      4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
          :|||::  |||: :::          ||:: : |: : |  : |         ||||
Db      7 LLLLAACAVPAQGLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 57

Qy     57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
          || : |::|     |:   :|: |||| |   |:
Db     58 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 86


RESULT   8
A41428
CEF-10 protein precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 03-Apr-1992 #sequence_revision 03-Apr-1992 #text_change 10-Sep-1997
C;Accession: A41428
R;Simmons, D.L.; Levy, D.B.; Yannoni, Y.; Erikson, R.L.
Proc. Natl. Acad. Sci. U.S.A. 86, 1178-1182, 1989
A;Title: Identification of a phorbol ester-repressible v-src-inducible gene.
A;Reference number: A41428; MUID:89145206
A;Accession: A41428
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-375 <SIM>
A;Cross-references: GB:J04496; NID:g211435; PID:g211436

  Query Match             13.5%;  Score 111;  DB 2;  Length 375;
  Best Local Similarity   29.1%;  Pred. No. 6.5e-05;
  Matches   39;  Conservative   19;  Mismatches   40;  Indels   36;  Gaps    9;

Qy      6 LLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRT---VLDDCGCCRVCAAGRG 62
          | ::||  |:|  : :      ||  |: :   ::|:|      | | ||||:|||
Db     10 LAAALLCLARLALGSP------CPAVCQCPA--AAPQCAPGVGLVPDGCGCCKVCAKQLN 61

Qy     63 ETCYRTVSGMDGMKCGP--GLRCQPSNGEDPFGEEFGICKD------CPYG--------T 106
          | | || :      |    || |  : |  | : : |||:       | |          :
Db     62 EDCSRTQP------CDHTKGLEC--NFGASPAATN-GICRAQSEGRPCEYNSKIYQNGES 112

Qy    107 FGMDCRETCNCQSG 120
          |  :|:  | |  |
Db    113 FQPNCKHQCTCIDG 126


RESULT   9
A53748
insulin-like growth factor-binding protein 5 precursor - human
N;Alternate names: IGFBP-5
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 31-Dec-1995 #text_change 20-Mar-1998
C;Accession: A53748; B40403; JH0391; PH0143
R;Allander, S.V.; Larsson, C.; Ehrenborg, E.; Suwanichkul, A.; Weber, G.; Morris, S.L
J. Biol. Chem. 269, 10891-10898, 1994
A;Title: Characterization of the chromosomal gene and promoter for human insulin-like
A;Reference number: A53748
A;Accession: A53748
A;Molecule type: DNA
A;Residues: 1-272 <ALL>
A;Cross-references: GB:L27556; GB:L27557; GB:L27558; GB:L27559; NID:g452057; PID:g505
R;Shimasaki, S.; Shimonaka, M.; Zhang, H.P.; Ling, N.
J. Biol. Chem. 266, 10646-10653, 1991
A;Title: Identification of five different insulin-like growth factor binding proteins
A;Reference number: A40403; MUID:91244847
A;Accession: B40403
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-272 <SHI>
A;Cross-references: GB:M62782; NID:g184817; PID:g184818
A;Experimental source: placenta
R;Kiefer, M.C.; Ioh, R.S.; Bauer, D.M.; Zapf, J.
Biochem. Biophys. Res. Commun. 176, 219-225, 1991
A;Title: Molecular cloning of a new human insulin-like growth factor binding protein.
A;Reference number: JH0391; MUID:91207396
A;Accession: JH0391
A;Molecule type: mRNA
A;Residues: 1-272 <KIE>
A;Cross-references: GB:M65062; NID:g184819; PID:g184820
A;Experimental source: osteosarcoma
R;Andress, D.L.; Birnbaum, R.S.
Biochem. Biophys. Res. Commun. 176, 213-218, 1991
A;Title: A novel human insulin-like growth factor binding protein secreted by osteobl
A;Reference number: PH0143; MUID:91207395
A;Accession: PH0143
A;Molecule type: protein
A;Residues: 24-37,'X',39-41,'X',43 <AND>
A;Experimental source: cell line V-2
C;Comment: Insulin-like growth factors occur in serum and other biological fluids com
C;Genetics:
A;Gene: GDB:IGFBP5
A;Cross-references: GDB:126837
A;Map position: 17q12-17q21
C;Superfamily: thyroglobulin type I repeat homology
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-272/Product: insulin-like growth factor-binding protein 5 #status predicted <IGF
F;192-263/Domain: thyroglobulin type I repeat homology <THY1>

  Query Match             13.4%;  Score 109.5;  DB 2;  Length 272;
```

```
  Best Local Similarity   27.7%;  Pred. No. 7.5e-05;
  Matches   26;  Conservative   20;  Mismatches   27;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::   ||: :::           ||:: : |: : |  : |          ||||
Db       8 LLLLAAYAGPAQSLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 58

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      59 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 87


RESULT  10
A40551
connective tissue growth factor - human
C;Species: Homo sapiens (man)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 20-Mar-1998
C;Accession: A40551; S44205
R;Bradham, D.M.; Igarashi, A.; Potter, R.L.; Grotendorst, G.R.
J. Cell Biol. 114, 1285-1294, 1991
A;Title: Connective tissue growth factor: a cysteine-rich mitogen secreted by human vasc
A;Reference number: A40551; MUID:91373462
A;Accession: A40551
A;Molecule type: mRNA
A;Residues: 1-349 <BRA>
A;Cross-references: GB:M92934; GB:M36965; GB:S56201; NID:g180923; PID:g180924
R;Oemar, B.S.; Werner, A.; Yang, Z.; Garnier, J.M.; Gentz, R.; Luescher, T.F.
submitted to the EMBL Data Library, April 1994
A;Description: Differential cloning and expression of human connective tissue growth fac
A;Reference number: S44205
A;Accession: S44205
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-349 <OEM>
A;Cross-references: EMBL:X78947; NID:g474933; PID:g474934


  Query Match             13.3%;  Score 109;  DB 2;  Length 349;
  Best Local Similarity   30.4%;  Pred. No. 0.0001;
  Matches   34;  Conservative   14;  Mismatches   36;  Indels   28;  Gaps    7;

Qy      27 DCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGP--GL 81
           :|:  |  ::  ::|||      ||| |||||||||   || |       :   | |  ||
Db      28 NCSGPCRCPD-EPAPRCPAGVSLVLDGCGCCRVCAKQLGELC------TERDPCDPHKGL 80

Qy      82 RCQPSNGEDPFGEEFGICKD-----CPYG--------TFGMDCRETCNCQSG 120
             |: ::   |     |:|       | :|        :|   |:  | |  |
Db      81 FCDFGS---PANRKIGVCTAKDGAPCIFGGTVYRSGESFQSSCKYQCTCLDG 129


RESULT  11
I47031
insulin-like growth factor-binding protein-4 - sheep (fragment)
C;Species: Ovis sp. (sheep)
C;Date: 15-Oct-1996 #sequence_revision 15-Oct-1996 #text_change 10-Oct-1997
C;Accession: I47031
R;Carr, J.M.; Grant, P.A.; Francis, G.L.; Owens, J.A.; Wallace, J.C.; Walton, P.E.
J. Mol. Endocrinol. 13, 219-236, 1994
A;Title: Isolation and characterization of ovine IGFBP-4: protein purification and cDNA
A;Reference number: I47031; MUID:95151165
A;Accession: I47031
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-237 <CAR>
A;Cross-references: GB:S77394; NID:g944951; PID:g944952
C;Superfamily: thyroglobulin type I repeat homology
F;153-228/Domain: thyroglobulin type I repeat homology <THY1>


  Query Match             13.0%;  Score 107;  DB 2;  Length 237;
  Best Local Similarity   34.8%;  Pred. No. 0.00013;
  Matches   24;  Conservative   13;  Mismatches   24;  Indels    8;  Gaps    3;

Qy      25 AVDCPQHCDS--SECKSSPRCKRTVLD-DCGCCRVCAAGRGETCYRTVSGMDGMKCGPGL 81
           |: ||    :   : |:::  |   | :  ||||  || |:|  |      |:    ||:||
Db       3 AIHCPPCSEEKLARCRPPVGCEELVREPGCGCCATCALGKGMPC-----GVYTPDCGSGL 57

Qy      82 RCQPSNGED 90
           ||:|: | :
Db      58 RCHPPRGVE 66


RESULT  12
G01662
insulin-like growth factor binding protein-4 - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
C;Accession: G01662
R;Strong, D.D.
submitted to the EMBL Data Library, February 1995
A;Reference number: G08078
A;Accession: G01662
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-258 <STR>
A;Cross-references: EMBL:U20982; NID:g695253; PID:g695254
C;Genetics:
A;Gene: GDB:IGFBP4
A;Cross-references: GDB:126811; OMIM:146733
A;Map position: 17q12-17q21
A;Introns: 117/1; 169/3; 214/3


  Query Match             13.0%;  Score 106.5;  DB 2;  Length 258;
  Best Local Similarity   31.8%;  Pred. No. 0.00015;
  Matches   28;  Conservative   22;  Mismatches   29;  Indels    9;  Gaps    4;

Qy       6 LLTTLLVPAHLVAAWSNNYAVDCPQHCDS--SECKSSPRCKRTVLD-DCGCCRVCAAGRG 62
           |:::||::|     :: ::: |: ||    :   : |:::  |   | :  ||||  || | |
Db       6 LVAALLLAAGPGPSLGDE-AIHCPPCSEEKLARCRPPVGCEELVREAGCGCCATCALGLG 64

Qy      63 ETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           |     |:   :|||:|||| |: | :
Db      65 MPC-----GVYTPRCGSGLRCYPPRGVE 87


RESULT  13
B37252
insulin-like growth factor-binding protein 4 precursor - human
N;Alternate names: IGFBP-4; inhibitory insulin-like growth factor-binding protein; in
C;Species: Homo sapiens (man)
C;Date: 28-Mar-1991 #sequence_revision 28-Mar-1991 #text_change 13-Sep-1998
C;Accession: B37252; B39842; A36549; A60712; A54650; A49801; A34419
R;Shimasaki, S.; Uchiyama, F.; Shimonaka, M.; Ling, N.
Mol. Endocrinol. 4, 1451-1458, 1990
A;Title: Molecular cloning of the cDNAs encoding a novel insulin-like growth factor-b
A;Reference number: A37252; MUID:91133415
A;Accession: B37252
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-258 <SHI>
R;Kiefer, M.C.; Masiarz, F.R.; Bauer, D.M.; Zapf, J.
J. Biol. Chem. 266, 9043-9049, 1991
A;Title: Identification and molecular cloning of two new 30-kDa insulin-like growth f
A;Reference number: A39842; MUID:91225006
A;Accession: B39842
A;Molecule type: mRNA
A;Residues: 1-258 <KIE>
A;Cross-references: GB:M62403; NID:g184815; PID:g184816
R;LaTour, D.; Mohan, S.; Linkhart, T.A.; Baylink, D.J.; Strong, D.D.
Mol. Endocrinol. 4, 1806-1814, 1990
A;Title: Inhibitory insulin-like growth factor-binding protein: cloning, complete seq
A;Reference number: A36549; MUID:91186988
A;Accession: A36549
A;Molecule type: mRNA
A;Residues: 1-50,'A',52-197,'F',199-258 <LAT>
```

```
R;Perkel, V.S.; Mohan, S.; Baylink, D.J.; Linkhart, T.A.
J. Clin. Endocrinol. Metab. 71, 533-535, 1990
A;Title: An inhibitory insulin-like growth factor binding protein (In-IGFBP) from human
A;Reference number: A60712
A;Accession: A60712
A;Molecule type: protein
A;Residues: 22-26,'X',28-29,'X',31-35 <PER>
R;Mohan, S.; Baylink, D.J.
Growth Regul. 1, 110-118, 1991
A;Title: Evidence that the inhibition of TE85 human bone cell proliferation by agents wh
A;Reference number: A54650; MUID:93091814
A;Accession: A54650
A;Molecule type: protein
A;Residues: 22-29,'X',31-42 <MOH>
A;Experimental source: TE85 osteosarcoma cells
A;Note: sequence extracted from NCBI backbone (NCBIP:121076)
R;Culouscou, J.M.; Shoyab, M.
Cancer Res. 51, 2813-2819, 1991
A;Title: Purification of a colon cancer cell growth inhibitor and its identification as
A;Reference number: A49801
A;Accession: A49801
A;Molecule type: protein
A;Residues: 22-53 <CUL>
A;Experimental source: colon adenocarcinoma cells
R;Mohan, S.; Bautista, C.M.; Wergedal, J.; Baylink, D.J.
Proc. Natl. Acad. Sci. U.S.A. 86, 8338-8342, 1989
A;Title: Isolation of an inhibitory insulin-like growth factor (IGF) binding protein fro
A;Reference number: A34419; MUID:90046792
A;Accession: A34419
A;Molecule type: protein
A;Residues: 22-29,'E',31-32,'A',34-36 <MO2>
C;Genetics:
A;Gene: GDB:IGFBP4
A;Cross-references: GDB:126811; OMIM:146733
A;Map position: 17q12-17q21
C;Superfamily: thyroglobulin type I repeat homology
C;Keywords: glycoprotein
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-258/Product: insulin-like growth factor-binding protein 4 #status experimental <MAT
F;174-249/Domain: thyroglobulin type I repeat homology <THY1>
F;125/Binding site: carbohydrate (Asn) (covalent) #status experimental


  Query Match             13.0%;  Score 106.5;  DB 2;  Length 258;
  Best Local Similarity   31.8%;  Pred. No. 0.00015;
  Matches   28;  Conservative   22;  Mismatches   29;  Indels    9;  Gaps    4;

Qy       6 LLTTLLVPAHLVAAWSNNYAVDCPQHCDS--SECKSSPRCKRTVLD-DCGCCRVCAAGRG 62
           |:::||::|   :: ::: |: ||    :   : |:::  |    | :  ||||  || | |
Db       6 LVAALLLAAGPGPSLGDE-AIHCPPCSEEKLARCRPPVGCEELVREPGCGCCATCALGLG 64

Qy      63 ETCYRTVSGMDGMKCGPGLRCQPSNGED 90
             |     |:   :||:|||| |: | :
Db      65 MPC-----GVYTPRCGSGLRCYPPRGVE 87


RESULT  14
A35037
insulin-like growth factor-binding protein 3 homolog - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 20-Jul-1990 #sequence_revision 20-Jul-1990 #text_change 01-Dec-1995
C;Accession: A35037
R;Shimasaki, S.; Shimonaka, M.; Ui, M.; Inouye, S.; Shibata, F.; Ling, N.
J. Biol. Chem. 265, 2198-2202, 1990
A;Title: Structural characterization of a follicle-stimulating hormone action inhibitor
A;Reference number: A35037; MUID:90130475
A;Accession: A35037
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-266 <SHI>
A;Cross-references: GB:J05228
C;Superfamily: insulin-like growth factor binding protein 1; thyroglobulin type I repeat
F;188-260/Domain: thyroglobulin type I repeat homology <THY1>


  Query Match             12.7%;  Score 104.5;  DB 2;  Length 266;
  Best Local Similarity   31.3%;  Pred. No. 0.00026;
  Matches   26;  Conservative   18;  Mismatches   20;  Indels   19;  Gaps    4;

Qy      31 HCDSSECK---------SSPRCKRTVLD-DCGCCRVCAAGRGETCYRTVSGMDGMKCGPG 80
           :|:: : :          ::| |    | :  ||||  ||   |::|      |:   :||:|
Db      12 RCEPCDARALAQCAPPPAAPPCAELVREPGCGCCLTCALREGQAC-----GVYTERCGAG 66

Qy      81 LRCQPSNGE----DPFGEEFGIC 99
           |||||: ||    ::: :  |||
Db      67 LRCQPPPGEPRPLQALLDGRGIC 89


RESULT  15
JC4584
insulin-like growth factor binding protein-5 - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 10-Apr-1996 #sequence_revision 24-May-1996 #text_change 10-Oct-1997
C;Accession: JC4584; G23734
R;White, M.E.; Diao, R.; Hathaway, M.R.; Mickelson, J.; Dayton, W.R.
Biochem. Biophys. Res. Commun. 218, 248-253, 1996
A;Title: Molecular cloning and sequence analysis of the porcine insulin-like growth f
A;Reference number: JC4584
A;Accession: JC4584
A;Molecule type: mRNA
A;Residues: 1-271 <WHI>
A;Cross-references: GB:U41340; NID:g1173906; PID:g1173907
A;Experimental source: skeletal muscle
R;Shimasaki, S.; Gao, L.; Shimonaka, M.; Ling, N.
Mol. Endocrinol. 5, 938-948, 1991
A;Title: Isolation and molecular cloning of insulin-like growth factor-binding protei
A;Reference number: A23734; MUID:92049376
A;Accession: G23734
A;Molecule type: protein
A;Residues: 20-25,'X',27-28,'X',30-36,'X',38-39 <SHI>
C;Comment: This protein has essential roles in the regulation and coordination of ins
lays a role during myoblast proliferation and differentiation, and is important in th
C;Superfamily: thyroglobulin type I repeat homology
C;Keywords: differentiation; growth factor; skeletal muscle
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-271/Product: insulin-like growth factor binding protein-5 #status experimental <
F;191-262/Domain: thyroglobulin type I repeat homology <THY1>


  Query Match             12.7%;  Score 104.5;  DB 2;  Length 271;
  Best Local Similarity   27.7%;  Pred. No. 0.00026;
  Matches   26;  Conservative   20;  Mismatches   27;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::   ||: :::           ||:: : |: : |  : |          ||||
Db       7 LLLLAACAGPAQGLGSFV---------HCEPCDEKALSMCPPSPLGCELVKDPGCGCCMT 57

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      58 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 86


Search completed: May  4, 1999, 08:18:15
Job time: 6967 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 4, 1999, 09:49:33 ; Search time 27.23 Seconds
(without alignments)
121.070 Million cell updates/sec

Title: US-09-037-460-2_COPY_1_163
Perfect score: 820
Sequence: 1 MKSVLLLTTLLVPAHLVAAW.....NRFVSLTEHDMASGDGNIVR 163

Scoring table: PAM150

Searched: 162890 seqs, 20225328 residues

Database : A_Geneseq_34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	820	100.0	184	1	R98994	Vascular IBP-like
2	141.5	17.3	351	1	R31599	Chicken nov protei
3	124	15.1	476	1	W63698	Human secreted pro
4	123.5	15.1	425	1	W56778	Homo sapiens D8725
5	123.5	15.1	480	1	W22849	Osteoblast like ce
6	119.5	14.6	258	1	W37466	Inhibitory IGF bin
7	112	13.7	348	1	R25566	Beta-IG-M2. TGF-be
8	112	13.7	348	1	W35731	Murine Fispl2. Iso
9	110.5	13.5	271	1	R26994	Rat IGFBP-5. DNA e
10	105	12.8	75	1	R31601	Chicken nov protei
11	109.5	13.4	272	1	R25700	IGFBP6. Insulin-li
12	109.5	13.4	272	1	R26995	Human IGFBP-5. DNA
13	109.5	13.4	272	1	R55084	Human insulin-like
14	109.5	13.4	272	1	R95329	Insulin-like growt
15	109.5	13.4	272	1	W35572	Insulin like growt
16	109	13.3	250	1	W37946	Human connective t
17	110.5	13.5	381	1	W35957	Human monocyte mat
18	103.5	12.6	76	1	R31600	Chicken nov protei
19	109	13.3	347	1	W12694	Connective tissue
20	109	13.3	349	1	R79964	Connective tissue
21	109	13.3	349	1	W11302	Connective tissue
22	109	13.3	349	1	W09089	Human connective t
23	109	13.3	349	1	W62084	Human connective t
24	109	13.3	349	1	W81425	Connective tissue
25	108.5	13.2	381	1	W35730	Human cysteine ric
26	106.5	13.0	258	1	R22253	Sequence of insuli
27	106.5	13.0	258	1	R21688	Sequence of insuli
28	106	12.9	264	1	R89951	Insulin-like growt
29	104	12.7	375	1	R90919	Connective tissue
30	102.5	12.5	291	1	R89952	Insulin-like growt
31	103.5	12.6	379	1	R25565	Beta-IG-M1. TGF-be
32	101.5	12.4	264	1	R13443	FSH inhibiting pro
33	101.5	12.4	264	1	R89950	Recombinant insuli
34	101.5	12.4	264	1	W12343	Human insulin-like
35	101.5	12.4	264	1	W12344	Human insulin-like
36	101.5	12.4	291	1	P92300	Sequence of human
37	101.5	12.4	291	1	R05596	Somatomedin carrie
38	101.5	12.4	291	1	R89273	Insulin like growt
39	101.5	12.4	1122	1	R73954	Human tie tyrosine
40	101.5	12.4	1138	1	R39820	tie receptor kinas
41	103	12.6	1801	1	W50895	Rat laminin B2 cha
42	94.5	11.5	259	1	P91868	Recombinant IGF bi
43	92	11.2	466	1	R07447	Human laminin B1 c
44	94.5	11.5	1764	1	P91672	Primary amino acid
45	94.5	11.5	1776	1	W50894	Mouse laminin B1 c

ALIGNMENTS

RESULT 1
R98994
ID R98994 standard; Protein; 184 AA.
AC R98994;
DT 06-NOV-1996 (first entry)
DE Vascular IBP-like growth factor.
KW Vascular IBF-like growth factor; VIGF;
KW insulin-like growth factor binding protein; agonist; antagonist;
KW muscle wastage; osteoporosis; implant fixation; wound healing;
KW therapy; diagnosis.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1. .21
FT /label= Sig_peptide
PN WO9617931-A1.
PD 13-JUN-1996.
PF 09-DEC-1994; U14388.
PR 09-DEC-1994; WO-U14388.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Hastings GA, Rosen CA;
DR WPI; 96-287176/29.
DR N-PSDB; T34991.
PT Human vascular insulin-like growth factor binding protein-like
PT growth factor, and its nucleic acid sequence and (ant)agonists -
PT used, e.g. to treat muscle wasting diseases or aid implant fixation,
PT or limit excess connective tissue prodn. during wound healing.
PS Claim 14; Page 43-44; 61pp; English.
CC Human vascular insulin-like growth factor binding protein-like
CC growth factor (R98994), or VIGF, is a protein of primarily
CC vascular origin that is structurally related to the IBP and CCN
CC protein families. It can be expressed in e.g. E. coli, CHO or
CC insect host cells using a vector incorporating a cDNA clone
CC (T34991), or its derivative, obtd. from human umbilical
CC endothelial cells. It is useful therapeutically e.g. for
CC treating muscle wasting diseases or osteoporosis, or can be used
CC to detect diseases associated with under- or over-expression of VIGF,
CC or to screen for antagonists useful during wound healing.
SQ Sequence 184 AA;

Query Match 100.0%; Score 820; DB 1; Length 184;
Best Local Similarity 100.0%; Pred. No. 4.9e-80;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLDDCGCCRVCAAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLDDCGCCRVCAAG 60

Qy      61 RGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICKDCPYGTFGMDCRETCNCQSG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 RGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICKDCPYGTFGMDCRETCNCQSG 120

Qy     121 ICDRGTGKCLKFPFFQYSVTKSSNRFVSLTEHDMASGDGNIVR 163
           |||||||||||||||||||||||||||||||||||||||||||
Db     121 ICDRGTGKCLKFPFFQYSVTKSSNRFVSLTEHDMASGDGNIVR 163
```

RESULT 2
R31599
ID R31599 standard; Protein; 351 AA.
AC R31599;
DT 24-MAY-1993 (first entry)
DE Chicken nov protein.
KW avian nephroblastoma; avian myeloblastoma virus; IGF binding site;
KW insulin-like growth factor; Wilm's tumour.
OS Gallus domesticus.

```
FH   Key             Location/Qualifiers
FT   peptide         1. .24
FT                   /label= signal_peptide
FT                   /note= "only hydrophobic region of protein"
FT   binding_site    56. .63
FT                   /label= IGF-binding_site_motif
FT                   /note= "corresponds to GCGCCXXC consensus"
PN   WO9300430-A.
PD   07-JAN-1993.
PF   25-JUN-1992; F00589.
PR   25-JUN-1991; FR-007807.
PA   (CNRS ) CENT NAT RECH SCI.
PI   Martinerie C, Perbal B;
DR   WPI; 93-036377/04.
DR   N-PSDB; Q36031.
PT   Nucleotide sequences hybridising to regions of chicken nov gene -
PT   useful as probes for detecting complementary sequences to
PT   evaluate development and/or differentiation of tumours
PS   Claim 1; Fig 1; 67pp; French.
CC   This amino acid sequence was deduced from the nucleotide sequence
CC   of a chicken nov gene clone isolated from a gene bank prepared from
CC   chicken embryonic fibroblasts screened with a tumour-derived probe.
CC   The only hydrophobic region occurs within the putative signal
CC   peptide suggesting that the protein is secreted. The protein also
CC   contains the consensus motif of proteins which bind to insulin-like
CC   growth factors. It is known that the human IGFII gene is
CC   overexpressed in some Wilm's tumours and a similar deregulation of
CC   IGFII expression could be involved in nephroblastoma development.
CC   The deduced nov protein sequence contains 39 (non-clustered)
CC   cysteine residues.
SQ   Sequence   351 AA;

  Query Match             17.3%;  Score 141.5;  DB 1;  Length 351;
  Best Local Similarity   31.9%;  Pred. No. 7.4e-08;
  Matches   45;  Conservative   17;  Mismatches   38;  Indels   41;  Gaps    7;

Qy       3 SVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAA 59
           :||||  ||: :  |:::       ||: | :     :|||     :||| |||| |||
Db      10 PVLLLLLLLLRPCEVSGR----EAACPRPCGGRCPAEPPRCAPGVPAVLDGCGCCLVCAR 65

Qy      60 GRGETCYRTVSGMDGMKCGPGLRCQPSNG-------EDPFGEEFGICK-----DCPYG-- 106
            |||:            |:| | |: | |       ||  |   |||      :| :
Db      66 QRGES-----------CSPLLPCDESGGLYCDRGPED--GGGAGICMVLEGDNCVFDGM 111

Qy     106 ------TFGMDCRETCNCQSG 120
                 ||   |:  | |: |
Db     112 IYRNGETFQPSCKYQCTCRDG 132


RESULT   3
W63698
ID   W63698 standard; Protein; 476 AA.
AC   W63698;
DT   24-SEP-1998  (first entry)
DE   Human secreted protein 18.
KW   Secreted protein; human; cell proliferation; cytokine activity;
KW   tissue growth; cellular differentiation; regeneration; activin;
KW   inhibin; chemotactic; haemostatic; thrombolytic; tumour inhibition;
KW   anti-inflammatory activity; biomarker.
OS   Homo sapiens.
PN   WO9825959-A2.
PD   18-JUN-1998.
PF   11-DEC-1997; U22787.
PR   11-DEC-1996; US-032757.
PA   (CHIR ) CHIRON CORP.
PI   Escobedo J, Garcia P, Hu Q, Kothakota S, Williams LT;
DR   WPI; 98-348453/30.
DR   N-PSDB; V43618.
PT   Secreted human polypeptides - having cytokine, cell proliferation or
PT   differentiation, activin or inhibin, tumour inhibition or
PT   anti-inflammatory activities
PS   Claim 1; Pages 70-72; 78pp; English.
CC   This represents a human secreted protein. The specification provides
CC   secreted protein sequences (W63681 to W63699) encoded by the nucleic
CC   acid sequences shown in V43601 to V43619. The invention provides a
CC   method of identifying a secreted polypeptide which is modified by rough
CC   microsomes. The secreted proteins can be used in assays to determine
CC   biological activities, such as cytokine, cell proliferation, or cellular
CC   differentiation activities, tissue growth or regeneration, activin or
CC   inhibin activity, chemotactic or chemokinetic activity, haemostatic or
CC   thrombolytic activity, receptor/ligand activity, tumour inhibition, or
CC   anti-inflammatory activity.  The proteins can also be used as biomarkers,
CC   to identify tissues or cell types which express the proteins, or a stage-
CC   or disease-specific alteration in protein expression. They can be used
CC   in protein interaction assays, to identify ligands or binding proteins.
CC   Compounds which affect the biological activities of the secreted proteins
CC   or their ability to interact with specific ligands can be identified
CC   using the proteins in screening assays. The proteins and antibodies that
CC   bind specifically to the protein can also be used to design diagnostic
CC   tests and therapeutic compositions for diseases which may be associated
CC   with altered expression of these proteins. Fusion proteins comprising,
CC   e.g. signal sequences or transmembrane domains of the proteins can be
CC   used to target other protein domains to cellular membrane or they can
CC   be secreted extracellularly.
SQ   Sequence   476 AA;

  Query Match             15.1%;  Score 124;  DB 1;  Length 476;
  Best Local Similarity   40.0%;  Pred. No. 7e-06;
  Matches   34;  Conservative   13;  Mismatches   28;  Indels   10;  Gaps    4;

Qy       5 LLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRC---KRTVLDDCGCCRVCAAGR 61
             |||  ||||    :|   :    ||  |::: | : | |      |:| | |||||:|:
Db      19 LLL--LLVPVLWAGAEKLHTQPSCPAVCQPTRCPALPTCALGTTPVFDLCRCCRVCPAAE 76

Qy      62 GETCYRTVSGMDGMKCGPGLRC-QP 85
             | |    :| :|  |:|||:| ||
Db      77 REVC----GGAQGQPCAPGLQCLQP 97


RESULT   4
W56778
ID   W56778 standard; Protein; 425 AA.
AC   W56778;
DT   13-OCT-1998  (first entry)
DE   Homo sapiens D87258 sequence.
KW   PS-1; presenilin; presenilin-1; PSP-1; Alzheimer's disease;
KW   serine protease; neurodegeneration; predisposition; diagnosis;
KW   D87258.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 213
FT                   /label= Gly, Val
PN   EP-828003-A2.
PD   11-MAR-1998.
PF   26-AUG-1997; 306501.
PR   13-DEC-1996; US-032875.
PR   06-SEP-1996; US-025436.
PR   25-OCT-1996; US-027873.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
PI   Browne MJ, Clinkenbeard HE, Creasy CL, Karran EH,
PI   Livi GP, Southan CD;
DR   WPI; 98-161101/15.
DR   N-PSDB; V29540.
PT   Nucleic acids encoding human serum protease protein(s) - used for
PT   diagnosing pre-disposition to Alzheimer's disease, etc.
PS   Claim 22; Page 30-31; 65pp; English.
CC   The sequence is that encoded by cDNA clone D87528 which can be used
CC   to identify modulators of serine protease activity and also to diagnose
CC   a condition associated with lack of one of the serine proteases
CC   or a genetic predisposition to neurodegeneration in a patient,
CC   preferably predisposition to Alzheimer's disease.
SQ   Sequence   425 AA;
```

```
  Query Match             15.1%;  Score 123.5;  DB 1;  Length 425;
  Best Local Similarity   34.0%;  Pred. No. 7.1e-06;
  Matches   33;  Conservative   18;  Mismatches   25;  Indels   21;  Gaps    5;

Qy       4 VLLLTTLLVPAHLV-AAWSNNYAVDCPQHCDSSECKSSP------RCKRTVLDDCGCCRV 56
           :||| :   :|:|  |::|   |  ||::|::: | : |      | :    | |||| |
Db      12 LLLLLAAPASAQLSRAGRSAPLAAGCPDRCEPARCPPQPEHCEGGRAR----DACGCCEV 67

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFG 93
           |:|  | :|     |::   || ||:|       |||
Db      68 CGAPEGAAC-----GLQEGPCGEGLQCVV-----PFG 94


RESULT   5
W22849
ID   W22849 standard; Protein; 480 AA.
AC   W22849;
DT   16-SEP-1997  (first entry)
DE   Osteoblast like cell derived protein.
KW   Osteoblast like cell; prevention; treatment; disease;
KW   contraceptive.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1. .27
FT                   /label= sig_peptide
FT   peptide         28. .480
FT                   /label= mat_peptide
PN   J09107980-A.
PD   28-APR-1997.
PF   12-AUG-1996; 231415.
PR   17-AUG-1995; JP-233537.
PA   (NISB ) JAPAN TOBACCO INC.
DR   WPI; 97-292469/27.
DR   N-PSDB; T75444.
PT   DNA encoding osteoblast like cell derived protein - useful for
PT   treatment and prevention of various diseases and as contraceptive
PS   Claim 2; Pages 26-27; 42pp; Japanese.
CC   The present sequence is an osteoblast like cell derived protein,
CC   which may be used to prevent or treat various diseases, or as a
CC   contraceptive.
CC   mRNA was collected from an osteoblast like cell, and double
CC   stranded DNA synthesised. A 3'-directed cDNA library was prepared,
CC   and the base sequence of the DNA of each clone in the cDNA library
CC   determined. An osteoblast complete chain cDNA library was prepared,
CC   and the clone GS2422, which has a complete chain, isolated. GS2422
CC   protein was expressed and purified, and an antibody against GS2422
CC   prepared. The amino acid sequence of a purified specimen of GS2422
CC   protein was analysed. The GS2422 gene is expressed in various human
CC   organs.
SQ   Sequence   480 AA;


  Query Match             15.1%;  Score 123.5;  DB 1;  Length 480;
  Best Local Similarity   34.0%;  Pred. No. 8e-06;
  Matches   33;  Conservative   18;  Mismatches   25;  Indels   21;  Gaps    5;

Qy       4 VLLLTTLLVPAHLV-AAWSNNYAVDCPQHCDSSECKSSP------RCKRTVLDDCGCCRV 56
           :||| :   :|:|  |::|   |  ||::|::: | : |      | :    | |||| |
Db      12 LLLLLAAPASAQLSRAGRSAPLAAGCPDRCEPARCPPQPEHCEGGRAR----DACGCCEV 67

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFG 93
           |:|  | :|     |::   || ||:|       |||
Db      68 CGAPEGAAC-----GLQEGPCGEGLQCVV-----PFG 94


RESULT   6
W37466
ID   W37466 standard; Protein; 258 AA.
AC   W37466;
DT   28-MAY-1998  (first entry)
DE   Inhibitory IGF binding protein.
KW   Inhibitory insulin like growth factor binding protein; In-IGF-BP;
KW   insulin like growth factor; IGF; bone cell; bone tumour; inhibition.
OS   Homo sapiens.
PN   US5693754-A.
PD   02-DEC-1997.
PF   22-OCT-1992; 966121.
PR   03-JUL-1990; US-548388.
PR   22-OCT-1992; US-966121.
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
PI   Baylink DJ, Mohan S;
DR   WPI; 98-031813/03.
DR   N-PSDB; V00257.
PT   Inhibitory insulin like growth factor binding protein - useful to
PT   inhibit effects of insulin like growth factor on bone cells and in
PT   diagnostic assays
PS   Claim 1; Fig 1; 11pp; English.
CC   This amino acid sequence is the novel inhibitory insulin like growth
CC   factor binding protein (In-IGF-BP), which is 258 amino acid residues in
CC   length.  The In-IGF-BP can be used to modulate the effects of insulin
CC   like growth factor I or II (IGF) on bone cells.  It is used to treat IGF
CC   dependent bone tumours, and in a competitive binding assay to determine
CC   IGF in clinical samples.
SQ   Sequence   258 AA;


  Query Match             14.6%;  Score 119.5;  DB 1;  Length 258;
  Best Local Similarity   32.4%;  Pred. No. 1.2e-05;
  Matches   34;  Conservative   24;  Mismatches   30;  Indels   17;  Gaps    6;

Qy       6 LLTTLLVPAHLVAAWSNNYAVDCPQHCD---SSECKSSPRCKRTVLDDCGCCRVCAAGRG 62
           |:::||::|   :: ::: |: ||   :   :  |:::  |   | : ||||  || | |
Db       6 LVAALLLAAGPGPSLGDE-AIHCPPCSEEKLARGCRPPVGCEELVREACGCCATCALGLG 64

Qy      63 ETCYRTVSGMDGMKCGPGLRCQPSNG-EDPF-----GEEFGICKD 101
             |     |:   :||:|||| |: | | |:     ||  |:| :
Db      65 MPC-----GVYTPRCGSGLRCYPPRGVEKPLHTLMHGE--GVCME 102


RESULT   7
R25566
ID   R25566 standard; Protein; 348 AA.
AC   R25566;
DT   18-JAN-1993  (first entry)
DE   Beta-IG-M2.
KW   Transforming growth factor beta; induced; CEF-10; v-src; chicken;
KW   embryo; fibroblasts; TGF-beta.
OS   Mus musculus.
PN   EP-495674-A.
PD   22-JUL-1992.
PF   17-JAN-1992; 300429.
PR   18-JAN-1991; US-642991.
PR   10-JAN-1992; US-816270.
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
PI   Brunner AM, Chinn J, Neubauer MG, Purchio AF;
DR   WPI; 92-243508/30.
DR   N-PSDB; Q26422.
PT   TGF-beta induced gene family - encodes proteins involved in
PT   growth and differentiation effects of TGF-beta-1
PS   Claim 3; Fig 2; 35pp; English.
CC   The protein sequence was deduced from the DNA sequence obtd. by
CC   screening a cDNA library made from AKR-2B mouse cells induced with
CC   TGF-beta1 and cyclohexamide with two probes from untreated AKR-2B
CC   mRNA and AKR-2B mRNA from cells treated with cyclohexamide and TGF-
CC   beta1.  The proteins encoded by hybridising colonies (beta-IG-M1 and
CC   beta-IG-M2) contain 38 Cys residues and are induced by TGF-beta1.
CC   Beta-IG-M2 displays 50 percent homology to the CEF-10 protein
CC   induced by v-src in chicken embryo fibroblasts.  Residues 52-59
CC   of beta-IG-M2 conform to the GCGCCXXC motif reported in the
CC   amino half of insulin-like growth factor (IGF) binding proteins.
CC   The C-terminal Cys rich region of beta-IG-M1, -M2 and CEF-10 contain
CC   an amino acid sequence with strong homology to a motif found near the
CC   C-terminal of the malarial circumsporozoite (CS) protein, which is
CC   highly conserved among all species of malarial parasites sequenced
CC   to date (designated region II).  This motif is also found in
```

```
CC   other proteins which have cell adhesive properties that mediate
CC   cell-cell and cell-extracellular matrix interactions, such as
CC   properdin, thrombospondin, and TRAP. The proteins encoded by
CC   TGF-beta induced genes are likely to be involved in mediation of
CC   the biological effects of TGF-beta relating to cell growth and
CC   differentiation.  See also R25565.
SQ   Sequence  348 AA;


  Query Match            13.7%;  Score 112;  DB 1;  Length 348;
  Best Local Similarity  30.4%;  Pred. No. 9.8e-05;
  Matches   34;  Conservative   15;  Mismatches   35;  Indels   28;  Gaps    7;

Qy      27 DCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGP--GL 81
           ||: :|: :   ::|:|      ||| ||||||||   || |       :   | |  ||
Db      27 DCSAQCQCAA-EAAPHCPAGVSLVLDGCGCCRVCAKQLGELC------TERDPCDPHKGL 79

Qy      82 RCQPSNGEDPFGEEFGICKD-----CPYG--------TFGMDCRETCNCQSG 120
             |: ::   |     |:|       | :|        :|   |:  | |  |
Db      80 FCDFGS---PANRKIGVCTAKDGAPCVFGGSVYRSGESFQSSCKYQCTCLDG 128


RESULT   8
W35731
ID   W35731 standard; Protein; 348 AA.
AC   W35731;
DT   29-MAR-1998  (first entry)
DE   Murine Fisp12.
KW   Fisp12; cysteine rich protein; mouse; Cyr61;
KW   extracellular matrix signalling molecule; cell adhesion;
KW   cell migration; cell proliferation; angiogenesis; chrondrogenesis;
KW   oncogenesis.
OS   Mus musculus.
PN   WO9733995-A2.
PD   18-SEP-1997.
PF   14-MAR-1997; U04193.
PR   15-MAR-1996; US-013958.
PA   (MUNI-) MUNIN CORP.
PI   Lau LF;
DR   WPI; 97-470875/43.
DR   N-PSDB; T94700.
PT   Isolated and purified cysteine rich protein 61, Cyr61 - useful to
PT   modulate e.g. haematostasis, induce wound healing, promote organ
PT   regeneration etc
PS   Example 2; Page 115-116; 133pp; English.
CC   This protein sequence comprises murine Fisp12, an extracellular
CC   matrix signalling molecule (ECM) that exhibits structural
CC   similarity to Cyr61 (see W35730) and which, like Cyr61, influences
CC   cell adhesion, proliferation and migration.  The human orthologue
CC   of Fisp12 is connective tissue growth factor.  Fisp12
CC   polynucleotides (see T94700) can be used for the production
CC   of Fisp12 polypeptides by recombinant methods.  Polypeptide
CC   compositions are provided that comprise mammalian ECM signalling
CC   molecules, peptide fragments, inhibitory peptides capable of
CC   interacting with receptors for ECM signalling molecules, and
CC   antibody products.  Further provided are methods for using
CC   mammalian ECM signalling molecules to screen for, and/or modulate
CC   disorders associated with angiogenesis, chondrogenesis and
CC   oncogenesis; ex vivo methods for using ECM signalling molecules
CC   to prepare blood products are also provided.
SQ   Sequence  348 AA;


  Query Match            13.7%;  Score 112;  DB 1;  Length 348;
  Best Local Similarity  30.4%;  Pred. No. 9.8e-05;
  Matches   34;  Conservative   15;  Mismatches   35;  Indels   28;  Gaps    7;

Qy      27 DCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGP--GL 81
           ||: :|: :   ::|:|      ||| ||||||||   || |       :   | |  ||
Db      27 DCSAQCQCAA-EAAPHCPAGVSLVLDGCGCCRVCAKQLGELC------TERDPCDPHKGL 79

Qy      82 RCQPSNGEDPFGEEFGICKD-----CPYG--------TFGMDCRETCNCQSG 120
             |: ::   |     |:|       | :|        :|   |:  | |  |
Db      80 FCDFGS---PANRKIGVCTAKDGAPCVFGGSVYRSGESFQSSCKYQCTCLDG 128


RESULT   9
R26994
ID   R26994 standard; Protein; 271 AA.
AC   R26994;
DT   16-FEB-1993  (first entry)
DE   Rat IGFBP-5.
KW   rat insulin-like growth factor binding protein-5; IGF-I; IGF-II;
KW   breast cancer; bone cancer; modulating bone growth; purification;
KW   affinity columns; antibodies; diagnosis; testing; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   peptide         1. .19
FT                   /note= "signal peptide"
PN   WO9214834-A.
PD   03-SEP-1992.
PF   13-FEB-1992; U01196.
PR   14-FEB-1991; US-658410.
PA   (WHIT-) WHITTIER INST DIABETES & ENDOCRINOLOGY.
PI   Ling NC, Shimaski S;
DR   WPI; 92-316186/38.
DR   N-PSDB; Q28270.
PT   DNA encoding insulin-like growth factor binding protein - useful
PT   for treating breast and bone cancer and modulating bone growth
PS   Disclosure; Page 11; 42pp; English.
CC   This sequence represents insulin-like growth factor binding protein.
CC   (IGBF-5) It was deduced from the appropriate nucleotide sequence
CC   obtained as in Q28270. IGBF-5 can bind to both IGFs -I and -iI. It is
CC   useful for treating conditions caused by an overabundance of IGFs eg.
CC   certain breast cancers and bone cancers. It is useful for modulating
CC   bone growth and in affinity chromatography columns for purification of
CC   IGF -I and -II. Antibodies to it can be used in assays to detect levels
CC   of the protein in mammals esp. humans and to neutralise the effects of
CC   IGFBP-5, and are useful in diagnostic test kits.
SQ   Sequence  271 AA;


  Query Match            13.5%;  Score 110.5;  DB 1;  Length 271;
  Best Local Similarity  28.7%;  Pred. No. 0.00011;
  Matches   27;  Conservative   20;  Mismatches   26;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::  |||: :::           ||:: : |: : |  : |         ||||
Db       7 LLLLAACAVPAQGLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 57

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |    |:
Db      58 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 86


RESULT  10
R31601
ID   R31601 standard; Protein; 75 AA.
AC   R31601;
DT   24-MAY-1993  (first entry)
DE   Chicken nov protein fragment V.
KW   avian nephroblastoma; avian myeloblastoma virus;
KW   stringent hybridisation.
OS   Gallus domesticus.
PN   WO9300430-A.
PD   07-JAN-1993.
PF   25-JUN-1992; F00589.
PR   25-JUN-1991; FR-007807.
PA   (CNRS ) CENT NAT RECH SCI.
PI   Martinerie C,  Perbal B;
DR   WPI; 93-036377/04.
PT   Nucleotide sequences hybridising to regions of chicken nov gene -
PT   useful as probes for detecting complementary sequences to
PT   evaluate development and/or differentiation of tumours
PS   Claim 5; Page 28; 67pp; French.
CC   The chicken nov gene is stimulated in avian nephroblastoma induced
```

```
CC   by avian myeloblastoma virus but not in normal adult kidney. A
CC   1975bp cDNA sequence (Q36031) was isolated from a gene bank prepared
CC   from chicken embryonic fibroblasts screened with a tumour-derived
CC   probe. Nucleotide sequences which hybridise to Q36031 or specified
CC   sub-fragments of it, under stringent conditions (i.e. 50% formamide,
CC   5 x SCC), are claimed. The claimed sequences preferably encode a
CC   protein with amino acid sequence V (R31601).
SQ   Sequence   75 AA;


  Query Match            12.8%;  Score 105;  DB 1;  Length 75;
  Best Local Similarity  38.7%;  Pred. No. 0.00013;
  Matches   29;  Conservative   15;  Mismatches   21;  Indels   10;  Gaps    4;

Qy      28 CPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGPGLRCQ 84
           || :| :    ::| |    |:||| |:|| ||| |||:|    | ::    ::|| |:
Db       7 CPPQCPGRCPATPPTCAPGVRAVLDGCSCCLVCARQRGESC----SDLEPCDESSGLYCD 62

Qy      85 PSNGEDPFGEEFGIC 99
             |   || ::: |||
Db      63 RS--ADP-SNQTGIC 74


RESULT  11
R25700
ID   R25700 standard; Protein; 272 AA.
AC   R25700;
DT   20-JAN-1993  (first entry)
DE   IGFBP6.
KW   Insulin; like; growth factor; binding protein; BP-6; IGF; growth;
KW   regeneration; hypopituitarism; oseoporosis; anaemia; breast cancer;
KW   kidney cancer; diabetic retinopathy; purification.
OS   Synthetic.
PN   WO9212243-A.
PD   23-JUL-1992.
PF   02-JAN-1992; U00107.
PR   08-JAN-1991; US-638628.
PA   (CHIR ) CHIRON CORP.
PI   Kiefer MC;
DR   WPI; 92-268666/32.
DR   P-PDSB; Q26814.
PT   Insulin-like growth factor binding protein-6 - for treating
PT   hypopituitarism, osteoporosis, anaemia(s), cancer, etc.,
PT   stimulating growth and wound healing, also useful in diagnosis
PS   Claim 2; Fig 1; 65pp; English.
CC   The protein sequence of insulin-like growth factor binding protein
CC   6 (IGFBP6) was deduced from the cDNA sequence obtd. by screening a
CC   ZAPII/human osteosarcoma cDNA library with IGFBP probes. IGFBP6
CC   may be used alone, or with IGF to stimulate growth, tissue or
CC   organ regeneration or wound healing.  Also IGFBP6 has applications
CC   in the treatment and diagnosis of hypopituitarism, osteoporosis,
CC   anaemias, and disorders due to excessive prodn. of free IGF, e.g.
CC   breast or kidney cancer, diabetic retinopathy, and abnormal growth
CC   of tall subjects.  The binding protein can also be used to purify
CC   IGF e.g. by affinity chromatography.
SQ   Sequence   272 AA;


  Query Match            13.4%;  Score 109.5;  DB 1;  Length 272;
  Best Local Similarity  27.7%;  Pred. No. 0.00014;
  Matches   26;  Conservative   20;  Mismatches   27;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::   ||: :::           ||:: : |: : |  : |         ||||
Db       8 LLLLAAYAGPAQSLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 58

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      59 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 87


RESULT  12
R26995
ID   R26995 standard; Protein; 272 AA.
AC   R26995;
DT   16-FEB-1993  (first entry)
DE   Human IGFBP-5.
KW   insulin-like growth factor binding protein-5; IGF-I; IGF-II;
KW   breast cancer; bone cancer; modulating bone growth; purification;
KW   affinity columns; antibodies; diagnosis; testing; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   peptide         1. .19
FT                   /note= "signal peptide"
PN   WO9214834-A.
PD   03-SEP-1992.
PF   13-FEB-1992; U01196.
PR   14-FEB-1991; US-658410.
PA   (WHIT-) WHITTIER INST DIABETES & ENDOCRINOLOGY.
PI   Ling NC, Shimaski S;
DR   WPI; 92-316186/38.
DR   N-PSDB; Q28271.
PT   DNA encoding insulin-like growth factor binding protein - useful
PT   for treating breast and bone cancer and modulating bone growth
PS   Disclosure; Page 11; 42pp; English.
CC   This sequence represents insulin-like growth factor binding protein.
CC   (IGBF-5) It was deduced from the appropriate nucleotide sequence
CC   obtained as in Q28271. IGBF-5 can bind to both IGFs -I and -II. It is
CC   useful for treating conditions caused by an overabundance of IGFs eg.
CC   certain breast cancers and bone cancers. It is useful for modulating
CC   bone growth and in affinity chromatography columns for purification of
CC   IGF -I and -II. Antibodies to it can be used in assays to detect levels
CC   of the protein in mammals esp. humans and to neutralise the effects of
CC   IGFBP-5, and are useful in diagnostic test kits.
SQ   Sequence   272 AA;


  Query Match            13.4%;  Score 109.5;  DB 1;  Length 272;
  Best Local Similarity  27.7%;  Pred. No. 0.00014;
  Matches   26;  Conservative   20;  Mismatches   27;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::   ||: :::           ||:: : |: : |  : |         ||||
Db       8 LLLLAAYAGPAQSLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 58

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      59 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 87


RESULT  13
R55084
ID   R55084 standard; Protein; 272 AA.
AC   R55084;
DT   28-NOV-1994  (first entry)
DE   Human insulin-like growth factor binding protein-5.
KW   Insulin-like growth factor binding protein-5; IGFBP-5; hormone.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         24. .163
FT                   /note= "preferred truncated IGFBP-5, Claim 12"
FT   peptide         21. .189
FT                   /note= "preferred truncated IGFBP-5, Claim 13"
PN   WO9410207-A.
PD   11-MAY-1994.
PF   29-OCT-1993; U10462.
PR   04-NOV-1992; US-972142.
PA   (CHIR ) CHIRON CORP.
PI   Andress DL,  Kiefer MC;
DR   WPI; 94-167395/20.
DR   N-PSDB; Q65519.
PT   Truncated insulin-like growth factor binding protein - has
PT   reduced affinity for insulin-like growth factor, useful for
PT   stimulating bone cell growth and mitogenic activity
PS   Disclosure; Fig. 1; 56pp; English.
CC   Within this protein sequence are contained 2 preferred truncated
```

```
CC   IGFBP-5 proteins derived from human U-2 osteosarcoma cells.
CC   The truncated IGFBPs may be used for stimulating mitogenic
CC   activity, particularly for stimulating bone cell growth.  The
CC   IGFBP is preferably produced recombinantly by expression in a
CC   yeast or CHO host.
SQ   Sequence   272 AA;


  Query Match            13.4%;  Score 109.5;  DB 1;  Length 272;
  Best Local Similarity   27.7%;  Pred. No. 0.00014;
  Matches   26;  Conservative   20;  Mismatches   27;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::   ||: :::           ||:: : |: : |  : |         ||||
Db       8 LLLLAAYAGPAQSLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 58

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      59 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 87


RESULT  14
R95329
ID   R95329 standard; Protein; 272 AA.
AC   R95329;
DT   09-DEC-1996  (first entry)
DE   Insulin-like growth factor binding protein-5 (IGFBP-5).
KW   TNF-R1-DD; tumour necrosis factor receptor 1 death domain; inhibitor;
KW   P55; anti-inflammatory; autoimmune disease; graft versus host reaction;
KW   osteoporosis; cachexia; diabetes; sequence identity; IGFBP-5;
KW   insulin-like growth factor binding protein-5.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   protein         87. .272
FT                   /label= clone 20DD
PN   WO9612735-A1.
PD   02-MAY-1996.
PF   12-OCT-1995; U12724.
PR   19-OCT-1994; US-327514.
PR   19-JUN-1995; US-494440.
PR   26-SEP-1995; US-533901.
PA   (GEMY ) GENETICS INST INC.
PI   Chen J,  Graham J,  Lin L,  Schievella AR;
DR   WPI; 96-230551/23.
DR   N-PSDB; T15231.
PT   TNF receptor death domain ligand proteins and inhibitors of ligand
PT   binding  - for prevention and treatment of pref. anti-inflammatory
PT   conditions, e.g. auto-immune disease, graft versus host reaction
PT   osteoporosis, etc.
PS   Claim 15; Page 42-43; 83pp; English.
CC   The present sequence is that of insulin-like growth factor binding
CC   protein-5 (IGFBP-5). Based upon the amino acid sequence identity between
CC   IGFBP-5 and a tumour necrosis factor (TNF) receptor 1 (R1) death domain
CC   (DD) ligand (clone 20DD; R95328) it has been determined that IGFBP-5 and
CC   certain fragments of it, will exhibit TNF-R1-DD ligand binding activity.
CC   A yeast genetic selection method, the "interaction trap", was used to
CC   screen W138 cell cDNA libraries for proteins that interact with the DD
CC   of the P55 type 1 TNF-R. TNF-R1-DD ligands and their inhibitors, e.g.
CC   IGFBP-5, are useful in the prevention and treatment of anti-inflammatory
CC   conditions and other conditions such as cachexia, autoimmune disease,
CC   graft versus host reaction, osteoporosis, diabetes, etc. .
SQ   Sequence   272 AA;


  Query Match            13.4%;  Score 109.5;  DB 1;  Length 272;
  Best Local Similarity   27.7%;  Pred. No. 0.00014;
  Matches   26;  Conservative   20;  Mismatches   27;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::   ||: :::           ||:: : |: : |  : |         ||||
Db       8 LLLLAAYAGPAQSLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 58

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      59 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 87


RESULT  15
W35572
ID   W35572 standard; Protein; 272 AA.
AC   W35572;
DT   19-MAR-1998  (first entry)
DE   Insulin like growth factor 5 binding protein.
KW   Tumour necrosis factor receptor P55 type; TNF-R1-DD ligand protein;
KW   death domain; TNF-R1; inhibitor identification; TNF-induced condition;
KW   insulin-like growth factor binding protein-5; inflammatory condition;
KW   IGFBP-5; therapy.
OS   Homo sapiens.
PN   WO9730084-A1.
PD   21-AUG-1997.
PF   11-FEB-1997; U02146.
PR   15-AUG-1996; US-698551.
PR   15-FEB-1996; US-602228.
PA   (GEMY ) GENETICS INST INC.
PI   Chen J, Graham J, Lin L, Schievella AR;
DR   WPI; 97-424976/39.
DR   N-PSDB; T94634.
PT   Tumour necrosis factor receptor P55 type death domain ligand
PT   proteins - useful for preventing or ameliorating inflammatory
PT   conditions
PS   Claim 15; Page 45-46; 103pp; English.
CC   This sequence represents a protein sequence of the invention. This
CC   sequence is the insulin-like growth factor binding protein-5 (IGFBP-5)
CC   and is a tumour necrosis factor receptor P55 type (TNF-R1) death domain
CC   (DD) ligand protein. A host cell containing DNA encoding this sequence is
CC   used for the recombinant production of TNF-R1-DD. The TNF-R1-DD ligand
CC   protein can be used in a method to identify inhibitors of TNF-R DD
CC   binding. The TNF-R1-DD ligand protein, IGFBP-5 (has TNF-R1-DD ligand
CC   activity), or inhibitors of TNF-R1-DD ligand protein are capable of
CC   preventing or ameliorating an inflammatory condition, preferably by
CC   inhibiting TNF-R DD binding. Identification and isolation of ligands
CC   allows their effects upon TNF-R signal transduction and use as
CC   therapeutic agents for treatment of TNF-induced conditions to be
CC   examined.
SQ   Sequence   272 AA;


  Query Match            13.4%;  Score 109.5;  DB 1;  Length 272;
  Best Local Similarity   27.7%;  Pred. No. 0.00014;
  Matches   26;  Conservative   20;  Mismatches   27;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::   ||: :::           ||:: : |: : |  : |         ||||
Db       8 LLLLAAYAGPAQSLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 58

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      59 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 87


Search completed: May  4, 1999, 12:32:37
Job time: 9784 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 4, 1999, 06:23:18 ; Search time 17.06 Seconds
(without alignments)
218.140 Million cell updates/sec

Title: US-09-037-460-2
Perfect score: 918
Sequence: 1 MKSVLLLTTLLVPAHLVAAW.....EVVKENAAGSPVMRKWLNPR 184

Scoring table: PAM150

Searched: 162890 seqs, 20225328 residues

Database : A_Geneseq_34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	918	100.0	184	1	R98994	Vascular IBP-like
2	141.5	15.4	351	1	R31599	Chicken nov protei
3	124	13.5	476	1	W63698	Human secreted pro
4	123.5	13.5	425	1	W56778	Homo sapiens D8725
5	123.5	13.5	480	1	W22849	Osteoblast like ce
6	119.5	13.0	258	1	W37466	Inhibitory IGF bin
7	112	12.2	348	1	R25566	Beta-IG-M2. TGF-be
8	112	12.2	348	1	W35731	Murine Fisp12. Iso
9	110.5	12.0	271	1	R26994	Rat IGFBP-5. DNA e
10	105	11.4	75	1	R31601	Chicken nov protei
11	109.5	11.9	272	1	R25700	IGFBP6. Insulin-li
12	109.5	11.9	272	1	R26995	Human IGFBP-5. DNA
13	109.5	11.9	272	1	R55084	Human insulin-like
14	109.5	11.9	272	1	R95329	Insulin-like growt
15	109.5	11.9	272	1	W35572	Insulin like growt
16	109	11.9	250	1	W37946	Human connective t
17	110.5	12.0	381	1	W35957	Human monocyte mat
18	103.5	11.3	76	1	R31600	Chicken nov protei
19	109	11.9	347	1	W12694	Connective tissue
20	109	11.9	349	1	R79964	Connective tissue
21	109	11.9	349	1	W11302	Connective tissue
22	109	11.9	349	1	W09089	Human connective t
23	109	11.9	349	1	W62084	Human connective t
24	109	11.9	349	1	W81425	Connective tissue
25	108.5	11.8	381	1	W35730	Human cysteine ric
26	106.5	11.6	258	1	R22253	Sequence of insuli
27	106.5	11.6	258	1	R21688	Sequence of insuli
28	106	11.5	264	1	R89951	Insulin-like growt
29	104	11.3	375	1	R90919	Connective tissue
30	102.5	11.2	291	1	R89952	Insulin-like growt
31	103.5	11.3	379	1	R25565	Beta-IG-M1. TGF-be
32	101.5	11.1	264	1	R13443	FSH inhibiting pro
33	101.5	11.1	264	1	R89950	Recombinant insuli
34	101.5	11.1	264	1	W12343	Human insulin-like
35	101.5	11.1	264	1	W12344	Human insulin-like
36	101.5	11.1	291	1	P92300	Sequence of human
37	101.5	11.1	291	1	R05596	Somatomedin carrie
38	101.5	11.1	291	1	R89273	Insulin like growt
39	101.5	11.1	1122	1	R73954	Human tie tyrosine
40	101.5	11.1	1138	1	R39820	tie receptor kinas
41	103	11.2	1801	1	W50895	Rat laminin B2 cha
42	94.5	10.3	259	1	P91868	Recombinant IGF bi
43	92	10.0	466	1	R07447	Human laminin B1 c
44	94.5	10.3	1764	1	P91672	Primary amino acid
45	94.5	10.3	1776	1	W50894	Mouse laminin B1 c

ALIGNMENTS

```
RESULT   1
R98994
ID   R98994 standard; Protein; 184 AA.
AC   R98994;
DT   06-NOV-1996  (first entry)
DE   Vascular IBP-like growth factor.
KW   Vascular IBF-like growth factor; VIGF;
KW   insulin-like growth factor binding protein; agonist; antagonist;
KW   muscle wastage; osteoporosis; implant fixation; wound healing;
KW   therapy; diagnosis.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         1. .21
FT                   /label= Sig_peptide
PN   WO9617931-A1.
PD   13-JUN-1996.
PF   09-DEC-1994; U14388.
PR   09-DEC-1994; WO-U14388.
PA   (HUMA-) HUMAN GENOME SCI INC.
PI   Hastings GA,  Rosen CA;
DR   WPI; 96-287176/29.
DR   N-PSDB; T34991.
PT   Human vascular insulin-like growth factor binding protein-like
PT   growth factor, and its nucleic acid sequence and (ant)agonists -
PT   used, e.g. to treat muscle wasting diseases or aid implant fixation,
PT   or limit excess connective tissue prodn. during wound healing.
PS   Claim 14; Page 43-44; 61pp; English.
CC   Human vascular insulin-like growth factor binding protein-like
CC   growth factor (R98994), or VIGF, is a protein of primarily
CC   vascular origin that is structurally related to the IBP and CCN
CC   protein families.  It can be expressed in e.g. E. coli, CHO or
CC   insect host cells using a vector incorporating a cDNA clone
CC   (T34991), or its derivative, obtd. from human umbilical
CC   endothelial cells.  It is useful therapeutically e.g. for
CC   treating muscle wasting diseases or osteoporosis, or can be used
CC   to detect diseases associated with under- or over-expression of VIGF,
CC   or to screen for antagonists useful during wound healing.
SQ   Sequence   184 AA;
```

```
  Query Match           100.0%;  Score 918;  DB 1;  Length 184;
  Best Local Similarity  100.0%;  Pred. No. 2.7e-94;
  Matches  184;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLDDCGCCRVCAAG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLDDCGCCRVCAAG 60

Qy      61 RGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICKDCPYGTFGMDCRETCNCQSG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 RGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICKDCPYGTFGMDCRETCNCQSG 120

Qy     121 ICDRGTGKCLKFPFFQYSVTKSSNRFVSLTEHDMASGDGNIVREEVVKENAAGSPVMRKW 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ICDRGTGKCLKFPFFQYSVTKSSNRFVSLTEHDMASGDGNIVREEVVKENAAGSPVMRKW 180

Qy     181 LNPR 184
           ||||
Db     181 LNPR 184
```

```
RESULT   2
R31599
ID   R31599 standard; Protein; 351 AA.
AC   R31599;
DT   24-MAY-1993  (first entry)
```

```
DE   Chicken nov protein.
KW   avian nephroblastoma; avian myeloblastoma virus; IGF binding site;
KW   insulin-like growth factor; Wilm's tumour.
OS   Gallus domesticus.
FH   Key             Location/Qualifiers
FT   peptide         1. .24
FT                   /label= signal_peptide
FT                   /note= "only hydrophobic region of protein"
FT   binding_site    56. .63
FT                   /label= IGF-binding_site_motif
FT                   /note= "corresponds to GCGCCXXC consensus"
PN   WO9300430-A.
PD   07-JAN-1993.
PF   25-JUN-1992; F00589.
PR   25-JUN-1991; FR-007807.
PA   (CNRS ) CENT NAT RECH SCI.
PI   Martinerie C,  Perbal B;
DR   WPI; 93-036377/04.
DR   N-PSDB; Q36031.
PT   Nucleotide sequences hybridising to regions of chicken nov gene -
PT   useful as probes for detecting complementary sequences to
PT   evaluate development and/or differentiation of tumours
PS   Claim 1; Fig 1; 67pp; French.
CC   This amino acid sequence was deduced from the nucleotide sequence
CC   of a chicken nov gene clone isolated from a gene bank prepared from
CC   chicken embryonic fibroblasts screened with a tumour-derived probe.
CC   The only hydrophobic region occurs within the putative signal
CC   peptide suggesting that the protein is secreted. The protein also
CC   contains the consensus motif of proteins which bind to insulin-like
CC   growth factors. It is known that the human IGFII gene is
CC   overexpressed in some Wilm's tumours and a similar deregulation of
CC   IGFII expression could be involved in nephroblastoma development.
CC   The deduced nov protein sequence contains 39 (non-clustered)
CC   cysteine residues.
SQ   Sequence   351 AA;
```

```
  Query Match             15.4%;  Score 141.5;  DB 1;  Length 351;
  Best Local Similarity   31.9%;  Pred. No. 3e-08;
  Matches   45;  Conservative   17;  Mismatches   38;  Indels   41;  Gaps    7;

Qy       3 SVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAA 59
           :||||  ||: :  |:::       ||: | :     :|||     :||| |||| |||
Db      10 PVLLLLLLLLRPCEVSGR----EAACPRPCGGRCPAEPPRCAPGVPAVLDGCGCCLVCAR 65

Qy      60 GRGETCYRTVSGMDGMKCGPGLRCQPSNG-------EDPFGEEFGICK-----DCPYG-- 106
           |||:            |:| | |: | |       || |   |||      :| :
Db      66 QRGES-----------CSPLLPCDESGGLYCDRGPED--GGGAGICMVLEGDNCVFDGM 111

Qy     106 ------TFGMDCRETCNCQSG 120
                 ||   |:  | |: |
Db     112 IYRNGETFQPSCKYQCTCRDG 132
```

```
RESULT   3
W63698
ID   W63698 standard; Protein; 476 AA.
AC   W63698;
DT   24-SEP-1998  (first entry)
DE   Human secreted protein 18.
KW   Secreted protein; human; cell proliferation; cytokine activity;
KW   tissue growth; cellular differentiation; regeneration; activin;
KW   inhibin; chemotactic; haemostatic; thrombolytic; tumour inhibition;
KW   anti-inflammatory activity; biomarker.
OS   Homo sapiens.
PN   WO9825959-A2.
PD   18-JUN-1998.
PF   11-DEC-1997; U22787.
PR   11-DEC-1996; US-032757.
PA   (CHIR ) CHIRON CORP.
PI   Escobedo J, Garcia P, Hu Q, Kothakota S, Williams LT;
DR   WPI; 98-348453/30.
DR   N-PSDB; V43618.
PT   Secreted human polypeptides - having cytokine, cell proliferation or
PT   differentiation, activin or inhibin, tumour inhibition or
PT   anti-inflammatory activities
PS   Claim 1; Pages 70-72; 78pp; English.
CC   This represents a human secreted protein. The specification provides
CC   secreted protein sequences (W63681 to W63699) encoded by the nucleic
CC   acid sequences shown in V43601 to V43619. The invention provides a
CC   method of identifying a secreted polypeptide which is modified by rough
CC   microsomes. The secreted proteins can be used in assays to determine
CC   biological activities, such as cytokine, cell proliferation, or cellular
CC   differentiation activities, tissue growth or regeneration, activin or
CC   inhibin activity, chemotactic or chemokinetic activity, haemostatic or
CC   thrombolytic activity, receptor/ligand activity, tumour inhibition, or
CC   anti-inflammatory activity.  The proteins can also be used as biomarkers,
CC   to identify tissues or cell types which express the proteins, or a stage-
CC   or disease-specific alteration in protein expression. They can be used
CC   in protein interaction assays, to identify ligands or binding proteins.
CC   Compounds which affect the biological activities of the secreted proteins
CC   or their ability to interact with specific ligands can be identified
CC   using the proteins in screening assays. The proteins and antibodies that
CC   bind specifically to the protein can also be used to design diagnostic
CC   tests and therapeutic compositions for diseases which may be associated
CC   with altered expression of these proteins. Fusion proteins comprising,
CC   e.g. signal sequences or transmembrane domains of the proteins can be
CC   used to target other protein domains to cellular membrane or they can
CC   be secreted extracellularly.
SQ   Sequence   476 AA;
```

```
  Query Match             13.5%;  Score 124;  DB 1;  Length 476;
  Best Local Similarity   40.0%;  Pred. No. 3.5e-06;
  Matches   34;  Conservative   13;  Mismatches   28;  Indels   10;  Gaps    4;

Qy       5 LLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRC---KRTVLDDCGCCRVCAAGR 61
           |||  ||||    :|   :    ||  |::: | : | |      |:| | |||||:|:
Db      19 LLL--LLVPVLWAGAEKLHTQPSCPAVCQPTRCPALPTCALGTTPVFDLCRCCRVCPAAE 76

Qy      62 GETCYRTVSGMDGMKCGPGLRC-QP 85
             | |    :| :|  |:|||:| ||
Db      77 REVC----GGAQGQPCAPGLQCLQP 97
```

```
RESULT   4
W56778
ID   W56778 standard; Protein; 425 AA.
AC   W56778;
DT   13-OCT-1998  (first entry)
DE   Homo sapiens D87258 sequence.
KW   PS-1; presenilin; presenilin-1; PSP-1; Alzheimer's disease;
KW   serine protease; neurodegeneration; predisposition; diagnosis;
KW   D87258.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   Misc_difference 213
FT                   /label= Gly, Val
PN   EP-828003-A2.
PD   11-MAR-1998.
PF   26-AUG-1997; 306501.
PR   13-DEC-1996; US-032875.
PR   06-SEP-1996; US-025436.
PR   25-OCT-1996; US-027873.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
PI   Browne MJ, Clinkenbeard HE, Creasy CL, Karran EH,
PI   Livi GP, Southan CD;
DR   WPI; 98-161101/15.
DR   N-PSDB; V29540.
PT   Nucleic acids encoding human serum protease protein(s) - used for
PT   diagnosing pre-disposition to Alzheimer's disease, etc.
PS   Claim 22; Page 30-31; 65pp; English.
CC   The sequence is that encoded by cDNA clone D87528 which can be used
CC   to identify modulators of serine protease activity and also to diagnose
CC   a condition associated with lack of one of the serine proteases
```

CC or a genetic predisposition to neurodegeneration in a patient,
CC preferably predisposition to Alzheimer's disease.
SQ Sequence 425 AA;

Query Match 13.5%; Score 123.5; DB 1; Length 425;
Best Local Similarity 34.0%; Pred. No. 3.6e-06;
Matches 33; Conservative 18; Mismatches 25; Indels 21; Gaps 5;

```
Qy       4 VLLLTTLLVPAHLV-AAWSNNYAVDCPQHCDSSECKSSP------RCKRTVLDDCGCCRV 56
           :||| :   :|:|  |::|   |  ||:::|::: | : |      | :    | |||| |
Db      12 LLLLLAAPASAQLSRAGRSAPLAAGCPDRCEPARCPPQPEHCEGGRAR----DACGCCEV 67

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFG 93
           |:|  | :|     |::   || ||:|       |||
Db      68 CGAPEGAAC-----GLQEGPCGEGLQCVV-----PFG 94
```

RESULT 5
W22849
ID W22849 standard; Protein; 480 AA.
AC W22849;
DT 16-SEP-1997 (first entry)
DE Osteoblast like cell derived protein.
KW Osteoblast like cell; prevention; treatment; disease;
KW contraceptive.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1. .27
FT /label= sig_peptide
FT peptide 28. .480
FT /label= mat_peptide
PN J09107980-A.
PD 28-APR-1997.
PF 12-AUG-1996; 231415.
PR 17-AUG-1995; JP-233537.
PA (NISB) JAPAN TOBACCO INC.
DR WPI; 97-292469/27.
DR N-PSDB; T75444.
PT DNA encoding osteoblast like cell derived protein - useful for
PT treatment and prevention of various diseases and as contraceptive
PS Claim 2; Pages 26-27; 42pp; Japanese.
CC The present sequence is an osteoblast like cell derived protein,
CC which may be used to prevent or treat various diseases, or as a
CC contraceptive.
CC mRNA was collected from an osteoblast like cell, and double
CC stranded DNA synthesised. A 3'-directed cDNA library was prepared,
CC and the base sequence of the DNA of each clone in the cDNA library
CC determined. An osteoblast complete chain cDNA library was prepared,
CC and the clone GS2422, which has a complete chain, isolated. GS2422
CC protein was expressed and purified, and an antibody against GS2422
CC prepared. The amino acid sequence of a purified specimen of GS2422
CC protein was analysed. The GS2422 gene is expressed in various human
CC organs.
SQ Sequence 480 AA;

Query Match 13.5%; Score 123.5; DB 1; Length 480;
Best Local Similarity 34.0%; Pred. No. 4.1e-06;
Matches 33; Conservative 18; Mismatches 25; Indels 21; Gaps 5;

```
Qy       4 VLLLTTLLVPAHLV-AAWSNNYAVDCPQHCDSSECKSSP------RCKRTVLDDCGCCRV 56
           :||| :   :|:|  |::|   |  ||:::|::: | : |      | :    | |||| |
Db      12 LLLLLAAPASAQLSRAGRSAPLAAGCPDRCEPARCPPQPEHCEGGRAR----DACGCCEV 67

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFG 93
           |:|  | :|     |::   || ||:|       |||
Db      68 CGAPEGAAC-----GLQEGPCGEGLQCVV-----PFG 94
```

RESULT 6
W37466
ID W37466 standard; Protein; 258 AA.
AC W37466;
DT 28-MAY-1998 (first entry)
DE Inhibitory IGF binding protein.
KW Inhibitory insulin like growth factor binding protein; In-IGF-BP;
KW insulin like growth factor; IGF; bone cell; bone tumour; inhibition.
OS Homo sapiens.
PN US5693754-A.
PD 02-DEC-1997.
PF 22-OCT-1992; 966121.
PR 03-JUL-1990; US-548388.
PR 22-OCT-1992; US-966121.
PA (BOEF) BOEHRINGER MANNHEIM GMBH.
PI Baylink DJ, Mohan S;
DR WPI; 98-031813/03.
DR N-PSDB; V00257.
PT Inhibitory insulin like growth factor binding protein - useful to
PT inhibit effects of insulin like growth factor on bone cells and in
PT diagnostic assays
PS Claim 1; Fig 1; 11pp; English.
CC This amino acid sequence is the novel inhibitory insulin like growth
CC factor binding protein (In-IGF-BP), which is 258 amino acid residues in
CC length. The In-IGF-BP can be used to modulate the effects of insulin
CC like growth factor I or II (IGF) on bone cells. It is used to treat IGF
CC dependent bone tumours, and in a competitive binding assay to determine
CC IGF in clinical samples.
SQ Sequence 258 AA;

Query Match 13.0%; Score 119.5; DB 1; Length 258;
Best Local Similarity 32.4%; Pred. No. 6e-06;
Matches 34; Conservative 24; Mismatches 30; Indels 17; Gaps 6;

```
Qy       6 LLTTLLVPAHLVAAWSNNYAVDCPQHCD---SSECKSSPRCKRTVLDDCGCCRVCAAGRG 62
           |:::||::|   :: ::: |: ||   :   :  |:::  |   | : ||||  || | |
Db       6 LVAALLLAAGPGPSLGDE-AIHCPPCSEEKLARGCRPPVGCEELVREACGCCATCALGLG 64

Qy      63 ETCYRTVSGMDGMKCGPGLRCQPSNG-EDPF-----GEEFGICKD 101
             |     |:   :||:|||| |: | | |:      ||  |:| :
Db      65 MPC-----GVYTPRCGSGLRCYPPRGVEKPLHTLMHGE--GVCME 102
```

RESULT 7
R25566
ID R25566 standard; Protein; 348 AA.
AC R25566;
DT 18-JAN-1993 (first entry)
DE Beta-IG-M2.
KW Transforming growth factor beta; induced; CEF-10; v-src; chicken;
KW embryo; fibroblasts; TGF-beta.
OS Mus musculus.
PN EP-495674-A.
PD 22-JUL-1992.
PF 17-JAN-1992; 300429.
PR 18-JAN-1991; US-642991.
PR 10-JAN-1992; US-816270.
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
PI Brunner AM, Chinn J, Neubauer MG, Purchio AF;
DR WPI; 92-243508/30.
DR N-PSDB; Q26422.
PT TGF-beta induced gene family - encodes proteins involved in
PT growth and differentiation effects of TGF-beta-1
PS Claim 3; Fig 2; 35pp; English.
CC The protein sequence was deduced from the DNA sequence obtd. by
CC screening a cDNA library made from AKR-2B mouse cells induced with
CC TGF-beta1 and cyclohexamide with two probes from untreated AKR-2B
CC mRNA and AKR-2B mRNA from cells treated with cyclohexamide and TGF-
CC beta1. The proteins encoded by hybridising colonies (beta-IG-M1 and
CC beta-IG-M2) contain 38 Cys residues and are induced by TGF-beta1.
CC Beta-IG-M2 displays 50 percent homology to the CEF-10 protein
CC induced by v-src in chicken embryo fibroblasts. Residues 52-59
CC of beta-IG-M2 conform to the GCGCCXXC motif reported in the
CC amino half of insulin-like growth factor (IGF) binding proteins.
CC The C-terminal Cys rich region of beta-IG-M1, -M2 and CEF-10 contain

```
CC   an amino acid sequence with strong homology to a motif found near the
CC   C-terminal of the malarial circumsporozoite (CS) protein, which is
CC   highly conserved among all species of malarial parasites sequenced
CC   to date (designated region II).  This motif is also found in
CC   other proteins which have cell adhesive properties that mediate
CC   cell-cell and cell-extracellular matrix interactions, such as
CC   properdin, thrombospondin, and TRAP.  The proteins encoded by
CC   TGF-beta induced genes are likely to be involved in mediation of
CC   the biological effects of TGF-beta relating to cell growth and
CC   differentiation.  See also R25565.
SQ   Sequence  348 AA;

  Query Match            12.2%;  Score 112;  DB 1;  Length 348;
  Best Local Similarity  30.4%;  Pred. No. 5.4e-05;
  Matches   34;  Conservative   15;  Mismatches   35;  Indels   28;  Gaps    7;

Qy      27 DCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGP--GL 81
           ||: :|: :   ::|:|      ||| ||||||||   || |       :   | |  ||
Db      27 DCSAQCQCAA-EAAPHCPAGVSLVLDGCGCCRVCAKQLGELC------TERDPCDPHKGL 79

Qy      82 RCQPSNGEDPFGEEFGICKD-----CPYG--------TFGMDCRETCNCQSG 120
             |: ::   |     |:|       | :|        :|   |:  | |  |
Db      80 FCDFGS---PANRKIGVCTAKDGAPCVFGGSVYRSGESFQSSCKYQCTCLDG 128


RESULT   8
W35731
ID   W35731 standard; Protein; 348 AA.
AC   W35731;
DT   27-MAR-1998  (first entry)
DE   Murine Fispl2.
KW   Fispl2; cysteine rich protein; mouse; Cyr61;
KW   extracellular matrix signalling molecule; cell adhesion;
KW   cell migration; cell proliferation; angiogenesis; chrondrogenesis;
KW   oncogenesis.
OS   Mus musculus.
PN   WO9733995-A2.
PD   18-SEP-1997.
PF   14-MAR-1997; U04193.
PR   15-MAR-1996; US-013958.
PA   (MUNI-) MUNIN CORP.
PI   Lau LF;
DR   WPI; 97-470875/43.
DR   N-PSDB; T94700.
PT   Isolated and purified cysteine rich protein 61, Cyr61 - useful to
PT   modulate e.g. haematostasis, induce wound healing, promote organ
PT   regeneration etc
PS   Example 2; Page 115-116; 133pp; English.
CC   This protein sequence comprises murine Fispl2, an extracellular
CC   matrix signalling molecule (ECM) that exhibits structural
CC   similarity to Cyr61 (see W35730) and which, like Cyr61, influences
CC   cell adhesion, proliferation and migration.  The human orthologue
CC   of Fispl2 is connective tissue growth factor.  Fispl2
CC   polynucleotides (see T94700) can be used for the production
CC   of Fispl2 polypeptides by recombinant methods.  Polypeptide
CC   compositions are provided that comprise mammalian ECM signalling
CC   molecules, peptide fragments, inhibitory peptides capable of
CC   interacting with receptors for ECM signalling molecules, and
CC   antibody products.  Further provided are methods for using
CC   mammalian ECM signalling molecules to screen for, and/or modulate
CC   disorders associated with angiogenesis, chondrogenesis and
CC   oncogenesis; ex vivo methods for using ECM signalling molecules
CC   to prepare blood products are also provided.
SQ   Sequence  348 AA;

  Query Match            12.2%;  Score 112;  DB 1;  Length 348;
  Best Local Similarity  30.4%;  Pred. No. 5.4e-05;
  Matches   34;  Conservative   15;  Mismatches   35;  Indels   28;  Gaps    7;

Qy      27 DCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGP--GL 81
           ||: :|: :   ::|:|      ||| ||||||||   || |       :   | |  ||
Db      27 DCSAQCQCAA-EAAPHCPAGVSLVLDGCGCCRVCAKQLGELC------TERDPCDPHKGL 79

Qy      82 RCQPSNGEDPFGEEFGICKD-----CPYG--------TFGMDCRETCNCQSG 120
             |: ::   |     |:|       | :|        :|   |:  | |  |
Db      80 FCDFGS---PANRKIGVCTAKDGAPCVFGGSVYRSGESFQSSCKYQCTCLDG 128


RESULT   9
R26994
ID   R26994 standard; Protein; 271 AA.
AC   R26994;
DT   16-FEB-1993  (first entry)
DE   Rat IGFBP-5.
KW   rat insulin-like growth factor binding protein-5; IGF-I; IGF-II;
KW   breast cancer; bone cancer; modulating bone growth; purification;
KW   affinity columns; antibodies; diagnosis; testing; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   peptide         1. .19
FT                   /note= "signal peptide"
PN   WO9214834-A.
PD   03-SEP-1992.
PF   13-FEB-1992; U01196.
PR   14-FEB-1991; US-658410.
PA   (WHIT-) WHITTIER INST DIABETES & ENDOCRINOLOGY.
PI   Ling NC, Shimaski S;
DR   WPI; 92-316186/38.
DR   N-PSDB; Q28270.
PT   DNA encoding insulin-like growth factor binding protein - useful
PT   for treating breast and bone cancer and modulating bone growth
PS   Disclosure; Page 11; 42pp; English.
CC   This sequence represents insulin-like growth factor binding protein.
CC   (IGBF-5) It was deduced from the appropriate nucleotide sequence
CC   obtained as in Q28270. IGBF-5 can bind to both IGFs -I and -iI. It is
CC   useful for treating conditions caused by an overabundance of IGFs eg.
CC   certain breast cancers and bone cancers. It is useful for modulating
CC   bone growth and in affinity chromatography columns for purification of
CC   IGF -I and -II. Antibodies to it can be used in assays to detect levels
CC   of the protein in mammals esp. humans and to neutralise the effects of
CC   IGFBP-5, and are useful in diagnostic test kits.
SQ   Sequence  271 AA;

  Query Match            12.0%;  Score 110.5;  DB 1;  Length 271;
  Best Local Similarity  28.7%;  Pred. No. 6.2e-05;
  Matches   27;  Conservative   20;  Mismatches   26;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::  |||: :::           ||:: : |: : |  : |         ||||
Db       7 LLLLAACAVPAQGLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 57

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      58 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 86


RESULT  10
R31601
ID   R31601 standard; Protein; 75 AA.
AC   R31601;
DT   24-MAY-1993  (first entry)
DE   Chicken nov protein fragment V.
KW   avian nephroblastoma; avian myeloblastoma virus;
KW   stringent hybridisation.
OS   Gallus domesticus.
PN   WO9300430-A.
PD   07-JAN-1993.
PF   25-JUN-1992; F00589.
PR   25-JUN-1991; FR-007807.
PA   (CNRS ) CENT NAT RECH SCI.
PI   Martinerie C,  Perbal B;
DR   WPI; 93-036377/04.
PT   Nucleotide sequences hybridising to regions of chicken nov gene -
```

```
PT   useful as probes for detecting complementary sequences to
PT   evaluate development and/or differentiation of tumours
PS   Claim 5; Page 28; 67pp; French.
CC   The chicken nov gene is stimulated in avian nephroblastoma induced
CC   by avian myeloblastoma virus but not in normal adult kidney. A
CC   1975bp cDNA sequence (Q36031) was isolated from a gene bank prepared
CC   from chicken embryonic fibroblasts screened with a tumour-derived
CC   probe. Nucleotide sequences which hybridise to Q36031 or specified
CC   sub-fragments of it, under stringent conditions (i.e. 50% formamide,
CC   5 x SCC), are claimed. The claimed sequences preferably encode a
CC   protein with amino acid sequence V (R31601).
SQ   Sequence   75 AA;

  Query Match             11.4%;  Score 105;  DB 1;  Length 75;
  Best Local Similarity   38.7%;  Pred. No. 6.8e-05;
  Matches   29;  Conservative   15;  Mismatches   21;  Indels   10;  Gaps    4;

Qy      28 CPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGPGLRCQ 84
           || :| :    ::| |    |:||| |:|| |||  |||:|    | ::    ::|| |:
Db       7 CPPQCPGRCPATPPTCAPGVRAVLDGCSCCLVCARQRGESC----SDLEPCDESSGLYCD 62

Qy      85 PSNGEDPFGEEFGIC 99
             |   || ::: |||
Db      63 RS--ADP-SNQTGIC 74


RESULT  11
R25700
ID   R25700 standard; Protein; 272 AA.
AC   R25700;
DT   20-JAN-1993  (first entry)
DE   IGFBP6.
KW   Insulin; like; growth factor; binding protein; BP-6; IGF; growth;
KW   regeneration; hypopituitarism; oseoporosis; anaemia; breast cancer;
KW   kidney cancer; diabetic retinopathy; purification.
OS   Synthetic.
PN   WO9212243-A.
PD   23-JUL-1992.
PF   02-JAN-1992; U00107.
PR   08-JAN-1991; US-638628.
PA   (CHIR ) CHIRON CORP.
PI   Kiefer MC;
DR   WPI; 92-268666/32.
DR   P-PDSB; Q26814.
PT   Insulin-like growth factor binding protein-6 - for treating
PT   hypopituitarism, osteoporosis, anaemia(s), cancer, etc.,
PT   stimulating growth and wound healing, also useful in diagnosis
PS   Claim 2; Fig 1; 65pp; English.
CC   The protein sequence of insulin-like growth factor binding protein
CC   6 (IGFBP6) was deduced from the cDNA sequence obtd. by screening a
CC   ZAPII/human osteosarcoma cDNA library with IGFBP probes. IGFBP6
CC   may be used alone, or with IGF to stimulate growth, tissue or
CC   organ regeneration or wound healing.  Also IGFBP6 has applications
CC   in the treatment and diagnosis of hypopituitarism, osteoporosis,
CC   anaemias, and disorders due to excessive prodn. of free IGF, e.g.
CC   breast or kidney cancer, diabetic retinopathy, and abnormal growth
CC   of tall subjects.  The binding protein can also be used to purify
CC   IGF e.g. by affinity chromatography.
SQ   Sequence   272 AA;

  Query Match             11.9%;  Score 109.5;  DB 1;  Length 272;
  Best Local Similarity   27.7%;  Pred. No. 8e-05;
  Matches   26;  Conservative   20;  Mismatches   27;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::   ||: :::           ||:: : |: : |  : |           ||||
Db       8 LLLLAAYAGPAQSLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 58

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: ||||  |   |:
Db      59 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 87


RESULT  12
R26995
ID   R26995 standard; Protein; 272 AA.
AC   R26995;
DT   16-FEB-1993  (first entry)
DE   Human IGFBP-5.
KW   insulin-like growth factor binding protein-5; IGF-I; IGF-II;
KW   breast cancer; bone cancer; modulating bone growth; purification;
KW   affinity columns; antibodies; diagnosis; testing; ss.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   peptide         1. .19
FT                   /note= "signal peptide"
PN   WO9214834-A.
PD   03-SEP-1992.
PF   13-FEB-1992; U01196.
PR   14-FEB-1991; US-658410.
PA   (WHIT-) WHITTIER INST DIABETES & ENDOCRINOLOGY.
PI   Ling NC, Shimaski S;
DR   WPI; 92-316186/38.
DR   N-PSDB; Q28271.
PT   DNA encoding insulin-like growth factor binding protein - useful
PT   for treating breast and bone cancer and modulating bone growth
PS   Disclosure; Page 11; 42pp; English.
CC   This sequence represents insulin-like growth factor binding protein.
CC   (IGBF-5) It was deduced from the appropriate nucleotide sequence
CC   obtained as in Q28271. IGBF-5 can bind to both IGFs -I and -II. It is
CC   useful for treating conditions caused by an overabundance of IGFs eg.
CC   certain breast cancers and bone cancers. It is useful for modulating
CC   bone growth and in affinity chromatography columns for purification of
CC   IGF -I and -II. Antibodies to it can be used in assays to detect levels
CC   of the protein in mammals esp. humans and to neutralise the effects of
CC   IGFBP-5, and are useful in diagnostic test kits.
SQ   Sequence   272 AA;

  Query Match             11.9%;  Score 109.5;  DB 1;  Length 272;
  Best Local Similarity   27.7%;  Pred. No. 8e-05;
  Matches   26;  Conservative   20;  Mismatches   27;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::   ||: :::           ||:: : |: : |  : |           ||||
Db       8 LLLLAAYAGPAQSLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 58

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: ||||  |   |:
Db      59 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 87


RESULT  13
R55084
ID   R55084 standard; Protein; 272 AA.
AC   R55084;
DT   28-NOV-1994  (first entry)
DE   Human insulin-like growth factor binding protein-5.
KW   Insulin-like growth factor binding protein-5; IGFBP-5; hormone.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   peptide         24. .163
FT                   /note= "preferred truncated IGFBP-5, Claim 12"
FT   peptide         21. .189
FT                   /note= "preferred truncated IGFBP-5, Claim 13"
PN   WO9410207-A.
PD   11-MAY-1994.
PF   29-OCT-1993; U10462.
PR   04-NOV-1992; US-972142.
PA   (CHIR ) CHIRON CORP.
PI   Andress DL,  Kiefer MC;
DR   WPI; 94-167395/20.
DR   N-PSDB; Q65519.
PT   Truncated insulin-like growth factor binding protein - has
```

```
PT   reduced affinity for insulin-like growth factor, useful for
PT   stimulating bone cell growth and mitogenic activity
PS   Disclosure; Fig. 1; 56pp; English.
CC   Within this protein sequence are contained 2 preferred truncated
CC   IGFBP-5 proteins derived from human U-2 osteosarcoma cells.
CC   The truncated IGFBPs may be used for stimulating mitogenic
CC   activity, particularly for stimulating bone cell growth.  The
CC   IGFBP is preferably produced recombinantly by expression in a
CC   yeast or CHO host.
SQ   Sequence   272 AA;

  Query Match             11.9%;  Score 109.5;  DB 1;  Length 272;
  Best Local Similarity   27.7%;  Pred. No. 8e-05;
  Matches   26;  Conservative   20;  Mismatches   27;  Indels   21;  Gaps    3;

Qy      4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
          :|||::   ||: :::          ||:: : |: : |  : |          ||||
Db      8 LLLLAAYAGPAQSLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 58

Qy     57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
          || : |::|     |:   :|: |||| |   |:
Db     59 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 87


RESULT  14
R95329
ID   R95329 standard; Protein; 272 AA.
AC   R95329;
DT   09-DEC-1996  (first entry)
DE   Insulin-like growth factor binding protein-5 (IGFBP-5).
KW   TNF-R1-DD; tumour necrosis factor receptor 1 death domain; inhibitor;
KW   P55; anti-inflammatory; autoimmune disease; graft versus host reaction;
KW   osteoporosis; cachexia; diabetes; sequence identity; IGFBP-5;
KW   insulin-like growth factor binding protein-5.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   protein         87. .272
FT                   /label= clone 20DD
PN   WO9612735-A1.
PD   02-MAY-1996.
PF   12-OCT-1995; U12724.
PR   19-OCT-1994; US-327514.
PR   19-JUN-1995; US-494440.
PR   26-SEP-1995; US-533901.
PA   (GEMY ) GENETICS INST INC.
PI   Chen J,  Graham J,  Lin L,  Schievella AR;
DR   WPI; 96-230551/23.
DR   N-PSDB; T15231.
PT   TNF receptor death domain ligand proteins and inhibitors of ligand
PT   binding  - for prevention and treatment of pref. anti-inflammatory
PT   conditions, e.g. auto-immune disease, graft versus host reaction
PT   osteoporosis, etc.
PS   Claim 15; Page 42-43; 83pp; English.
CC   The present sequence is that of insulin-like growth factor binding
CC   protein-5 (IGFBP-5). Based upon the amino acid sequence identity between
CC   IGFBP-5 and a tumour necrosis factor (TNF) receptor 1 (R1) death domain
CC   (DD) ligand (clone 20DD; R95328) it has been determined that IGFBP-5 and
CC   certain fragments of it, will exhibit TNF-R1-DD ligand binding activity.
CC   A yeast genetic selection method, the "interaction trap", was used to
CC   screen W138 cell cDNA libraries for proteins that interact with the DD
CC   of the P55 type 1 TNF-R. TNF-R1-DD ligands and their inhibitors, e.g.
CC   IGFBP-5, are useful in the prevention and treatment of anti-inflammatory
CC   conditions and other conditions such as cachexia, autoimmune disease,
CC   graft versus host reaction, osteoporosis, diabetes, etc. .
SQ   Sequence   272 AA;

  Query Match             11.9%;  Score 109.5;  DB 1;  Length 272;
  Best Local Similarity   27.7%;  Pred. No. 8e-05;
  Matches   26;  Conservative   20;  Mismatches   27;  Indels   21;  Gaps    3;

Qy      4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
          :|||::   ||: :::          ||:: : |: : |  : |          ||||
Db      8 LLLLAAYAGPAQSLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 58

Qy     57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
          || : |::|     |:   :|: |||| |   |:
Db     59 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 87


RESULT  15
W35572
ID   W35572 standard; Protein; 272 AA.
AC   W35572;
DT   19-MAR-1998  (first entry)
DE   Insulin like growth factor 5 binding protein.
KW   Tumour necrosis factor receptor P55 type; TNF-R1-DD ligand protein;
KW   death domain; TNF-R1; inhibitor identification; TNF-induced condition;
KW   insulin-like growth factor binding protein-5; inflammatory condition;
KW   IGFBP-5; therapy.
OS   Homo sapiens.
PN   WO9730084-A1.
PD   21-AUG-1997.
PF   11-FEB-1997; U02146.
PR   15-AUG-1996; US-698551.
PR   15-FEB-1996; US-602228.
PA   (GEMY ) GENETICS INST INC.
PI   Chen J, Graham J, Lin L, Schievella AR;
DR   WPI; 97-424976/39.
DR   N-PSDB; T94634.
PT   Tumour necrosis factor receptor P55 type death domain ligand
PT   proteins - useful for preventing or ameliorating inflammatory
PT   conditions
PS   Claim 15; Page 45-46; 103pp; English.
CC   This sequence represents a protein sequence of the invention. This
CC   sequence is the insulin-like growth factor binding protein-5 (IGFBP-5)
CC   and is a tumour necrosis factor receptor P55 type (TNF-R1) death domain
CC   (DD) ligand protein. A host cell containing DNA encoding this sequence is
CC   used for the recombinant production of TNF-R1-DD. The TNF-R1-DD ligand
CC   protein can be used in a method to identify inhibitors of TNF-R DD
CC   binding. The TNF-R1-DD ligand protein, IGFBP-5 (has TNF-R1-DD ligand
CC   activity), or inhibitors of TNF-R1-DD ligand protein are capable of
CC   preventing or ameliorating an inflammatory condition, preferably by
CC   inhibiting TNF-R DD binding. Identification and isolation of ligands
CC   allows their effects upon TNF-R signal transduction and use as
CC   therapeutic agents for treatment of TNF-induced conditions to be
CC   examined.
SQ   Sequence   272 AA;

  Query Match             11.9%;  Score 109.5;  DB 1;  Length 272;
  Best Local Similarity   27.7%;  Pred. No. 8e-05;
  Matches   26;  Conservative   20;  Mismatches   27;  Indels   21;  Gaps    3;

Qy      4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
          :|||::   ||: :::          ||:: : |: : |  : |          ||||
Db      8 LLLLAAYAGPAQSLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 58

Qy     57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
          || : |::|     |:   :|: |||| |   |:
Db     59 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 87


Search completed: May  4, 1999, 08:18:43
Job time: 6925 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 4, 1999, 13:59:22 ; Search time 11.8 Seconds
(without alignments)
418.527 Million cell updates/sec

Title: US-09-037-460-2
Perfect score: 918
Sequence: 1 MKSVLLLTTLLVPAHLVAAW.....EVVKENAAGSPVMRKWLNPR 184

Scoring table: PAM150

Searched: 74019 seqs, 26840295 residues

Database : SwissProt_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	141.5	15.4	351	1	NOV_CHICK	P28686 gallus gall
2	139	15.1	353	1	NOV_COTJA	P42642 coturnix co
3	124.5	13.6	354	1	NOV_MOUSE	Q64299 mus musculu
4	123.5	13.5	357	1	NOV_HUMAN	P48745 homo sapien
5	115.5	12.6	258	1	IBP4_BOVIN	Q05716 bos taurus
6	114.5	12.5	271	1	IBP5_MOUSE	Q07079 mus musculu
7	112	12.2	348	1	CTGF_MOUSE	P29268 mus musculu
8	110.5	12.0	271	1	IBP5_RAT	P24594 rattus norv
9	111	12.1	375	1	CE10_CHICK	P19336 gallus gall
10	109.5	11.9	272	1	IBP5_HUMAN	P24593 homo sapien
11	110.5	12.0	381	1	CYR6_HUMAN	O00622 homo sapien
12	109	11.9	349	1	CTGF_HUMAN	P29279 homo sapien
13	107	11.7	237	1	IBP4_SHEEP	Q28893 ovis aries
14	106.5	11.6	258	1	IBP4_HUMAN	P22692 homo sapien
15	104.5	11.4	266	1	IBP3_PIG	P16611 sus scrofa
16	104.5	11.4	271	1	IBP5_PIG	Q28985 sus scrofa
17	103.5	11.3	254	1	IBP4_MOUSE	P47879 mus musculu
18	103.5	11.3	254	1	IBP4_RAT	P21744 rattus norv
19	103.5	11.3	379	1	CYR6_MOUSE	P18406 mus musculu
20	103	11.2	349	1	CTGF_PIG	O19113 sus scrofa
21	101.5	11.1	291	1	IBP3_BOVIN	P20959 bos taurus
22	101.5	11.1	291	1	IBP3_HUMAN	P17936 homo sapien
23	100.5	10.9	291	1	IBP3_MOUSE	P47878 mus musculu
24	99	10.8	343	1	NOV_XENLA	P51609 xenopus lae
25	98.5	10.7	292	1	IBP3_RAT	P15473 rattus norv
26	103	11.2	1801	1	LMB2_RAT	P15800 rattus norv
27	101.5	11.1	1138	1	TIE1_HUMAN	P35590 homo sapien
28	97	10.6	317	1	IBP2_SHEEP	Q29400 ovis aries
29	94.5	10.3	259	1	IBP1_HUMAN	P08833 homo sapien
30	100	10.9	1799	1	LMB2_MOUSE	Q61292 mus musculu
31	94	10.2	311	1	IBP2_CHICK	P49705 gallus gall
32	97.5	10.6	1136	1	TIE1_BOVIN	Q06805 bos taurus
33	87.5	9.5	111	1	IBP5_BOVIN	Q05717 bos taurus
34	94.5	10.3	1134	1	TIE1_MOUSE	Q06806 mus musculu
35	94.5	10.3	1786	1	LMB1_MOUSE	P02469 mus musculu
36	95.5	10.4	3635	1	LMA5_MOUSE	Q61001 mus musculu
37	92	10.0	1786	1	LMB1_HUMAN	P07942 homo sapien
38	86.5	9.4	328	1	IBP2_HUMAN	P18065 homo sapien
39	91.5	10.0	1816	1	LMA4_HUMAN	Q16363 homo sapien
40	86	9.4	349	1	CTGF_BOVIN	O18739 bos taurus
41	94	10.2	5147	1	FAT_DROME	P33450 drosophila
42	90	9.8	1639	1	LMG1_DROME	P15215 drosophila
43	84.5	9.2	304	1	IBP2_RAT	P12843 rattus norv
44	84.5	9.2	305	1	IBP2_MOUSE	P47877 mus musculu
45	83.5	9.1	249	1	YN85_YEAST	P53718 saccharomyc

ALIGNMENTS

```
RESULT   1
NOV_CHICK
ID   NOV_CHICK       STANDARD;       PRT;   351 AA.
AC   P28686;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   NOV PROTEIN PRECURSOR.
GN   NOV.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC   GALLIFORMES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BROWN LEGHORN;
RX   MEDLINE; 92107157.
RA   MALOISEL V., MARTINERIE C., DAMBRINE G., PLASSIART G., BRISAC M.,
RA   CROCHET J., PERBAL B.;
RL   MOL. CELL. BIOL. 12:10-21(1992).
CC   -!- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL
CC       GROWTH REGULATION. ITS OVEREXPRESSION IS ASSOCIATED WITH
CC       TUMORIGENESIS AND EXPRESSION OF A N-TERMINAL-TRUNCATED VERSION
CC       OF NOV GENE IN CHICKEN EMBRYONIC FIBROBLASTS (CEF) IS SUFFICIENT
CC       TO INDUCE THE TRANSFORMATION OF CEF IN VITRO.
CC   -!- TISSUE SPECIFICITY: BRAIN AND HEART, AND AT A LOWER LEVEL IN
CC       MUSCLE AND INTESTINE, IN THE EMBRYO. LUNG AND LESS SO IN BRAIN AND
CC       SPLEEN, IN ADULT CHICKEN.
CC   -!- DEVELOPMENTAL STAGE: MAV1-INDUCED NEPHROBLASTOMAS EXPRESS A HIGH
CC       LEVEL OF NOV GENE WHOSE TRANSCRIPTION IS NORMALLY ARRESTED IN
CC       ADULT KIDNEY.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; X59284; G63703; -.
DR   PIR; S20078; S20078.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   PROTO-ONCOGENE; GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    351       NOV PROTEIN.
FT   DOMAIN      104    170       VWFC.
FT   DOMAIN      258    332       CTCK.
FT   DISULFID    258    295       BY SIMILARITY.
FT   DISULFID    275    309       BY SIMILARITY.
FT   DISULFID    286    325       BY SIMILARITY.
FT   DISULFID    289    327       BY SIMILARITY.
FT   DISULFID    294    331       BY SIMILARITY.
FT   CARBOHYD    274    274       POTENTIAL.
SQ   SEQUENCE   351 AA;  38268 MW;  C7044065 CRC32;
```

Query Match 15.4%; Score 141.5; DB 1; Length 351;
Best Local Similarity 31.9%; Pred. No. 3.3e-08;
Matches 45; Conservative 17; Mismatches 38; Indels 41; Gaps 7;

```
Qy       3 SVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAA 59
           :||||  ||: :  |:::          ||: | :     :|||     :||| |||| |||
Db      10 PVLLLLLLLLRPCEVSGR----EAACPRPCGGRCPAEPPRCAPGVPAVLDGCGCCLVCAR 65

Qy      60 GRGETCYRTVSGMDGMKCGPGLRCQPSNG-------EDPFGEEFGICK-----DCPYG-- 106
           |||:             |:| | |: | |       ||  |  |||      :| :
Db      66 QRGES-----------CSPLLPCDESGGLYCDRGPED--GGGAGICMVLEGDNCVFDGM 111
```

```
Qy    106 ------TFGMDCRETCNCQSG 120
                ||   |:  | |: |
Db    112 IYRNGETFQPSCKYQCTCRDG 132


RESULT   2
NOV_COTJA
ID   NOV_COTJA      STANDARD;      PRT;   353 AA.
AC   P42642;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   NOV PROTEIN PRECURSOR.
GN   NOV.
OS   COTURNIX COTURNIX JAPONICA (JAPANESE QUAIL).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC   GALLIFORMES.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   WEISKIRCHEN R., BISTER K.;
RL   SUBMITTED (AUG-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL
CC       GROWTH REGULATION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; U13063; G532697; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   PROTO-ONCOGENE; GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27    353       NOV PROTEIN.
FT   DOMAIN      106    172       VWFC.
FT   DOMAIN      260    334       CTCK.
FT   DISULFID    260    297       BY SIMILARITY.
FT   DISULFID    277    311       BY SIMILARITY.
FT   DISULFID    288    327       BY SIMILARITY.
FT   DISULFID    291    329       BY SIMILARITY.
FT   DISULFID    296    333       BY SIMILARITY.
FT   CARBOHYD    276    276       POTENTIAL.
SQ   SEQUENCE   353 AA;  38667 MW;  C4F5928D CRC32;


  Query Match             15.1%;  Score 139;  DB 1;  Length 353;
  Best Local Similarity   32.1%;  Pred. No. 6.1e-08;
  Matches   45;  Conservative   14;  Mismatches   39;  Indels   42;  Gaps    7;

Qy      4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAG 60
          :|||  || |: :     |     ||: | :     :|||     :||| |||| |||
Db     14 LLLLLLLLRPSEV-----NGREAPCPRPCGGRCPAEPPRCAPGVPAVLDGCGCCLVCARQ 68

Qy     61 RGETCYRTVSGMDGMKCGPGLRCQPSNG-------EDPFGEEFGICK-----DCPYG--- 106
          |||:            |:| | |: | |       ||  |   |||      :| :
Db     69 RGES-----------CSPLLPCDESGGLYCDRGPED--GGGTGICMVLEGDNCVFDGMI 114

Qy    106 -----TFGMDCRETCNCQSG 120
               ||   |:  | |: |
Db    115 YRNGETFQPSCKYQCTCRDG 134


RESULT   3
NOV_MOUSE
ID   NOV_MOUSE      STANDARD;      PRT;   354 AA.
AC   Q64299;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   NOV PROTEIN HOMOLOG PRECURSOR (NOVH).
GN   NOV.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SV, AND ICR; TISSUE=BRAIN;
RA   SNAITH M.R., NATARAJAN D., TAYLOR L.B., CHOI C.P., MARTINERIE C.,
RA   PERBAL B., SCHOFIELD P.N., BOULTER C.A.;
RL   SUBMITTED (MAY-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE; 96204003.
RA   MARTINERIE C., CHEVALIER G., RAUSCHER F.J. III, PERBAL B.;
RL   ONCOGENE 12:1479-1492(1996).
CC   -!- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL
CC       GROWTH REGULATION (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; X97863; E1181580; -.
DR   EMBL; Y09257; E281106; -.
DR   EMBL; X96585; E228599; -.
DR   MGD; MGI:109185; NOV.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   PROTO-ONCOGENE; GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    354       NOV PROTEIN HOMOLOG.
FT   DOMAIN      102    168       VWFC.
FT   DOMAIN      261    335       CTCK.
FT   DISULFID    261    298       BY SIMILARITY.
FT   DISULFID    278    312       BY SIMILARITY.
FT   DISULFID    289    328       BY SIMILARITY.
FT   DISULFID    292    330       BY SIMILARITY.
FT   DISULFID    297    334       BY SIMILARITY.
FT   CARBOHYD     91     91       POTENTIAL.
FT   CARBOHYD    277    277       POTENTIAL.
SQ   SEQUENCE   354 AA;  38928 MW;  65A4FF0E CRC32;


  Query Match             13.6%;  Score 124.5;  DB 1;  Length 354;
  Best Local Similarity   32.7%;  Pred. No. 2.2e-06;
  Matches   37;  Conservative   18;  Mismatches   27;  Indels   31;  Gaps    9;

Qy     28 CPQHCDSSECKS-SPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMK-C--GPG 80
          || :| :: | | || |    |:||| |:|| |||  |||:|         |: |  ::|
Db     29 CPSRC-PPKCPSISPTCAPGVRSVLDGCSCCPVCARQRGESC-------SEMRPCDQSSG 80

Qy     81 LRCQPSNGEDPFGEEFGICK-----DCPYG--------TFGMDCRETCNCQSG 120
          | |: |   :   :: |||      :| :          |  :|:  | |: |
Db     81 LYCDRSADPN---NQTGICMVPEGDNCVFDGVIYRNGEKFEPNCQYFCTCRDG 130


RESULT   4
NOV_HUMAN
ID   NOV_HUMAN      STANDARD;      PRT;   357 AA.
AC   P48745;
DT   01-FEB-1996 (REL. 33, CREATED)
DT   01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   NOV PROTEIN HOMOLOG PRECURSOR (NOVH).
GN   NOV.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 94336229.
RA   MARTINERIE C., HUFF V., JOUBERT I., BADZIOCH M., SAUNDERS G.,
```

```
RA   STRONG L., PERBAL B.;
RL   ONCOGENE 9:2729-2732(1994).
CC   -!- FUNCTION: IMMEDIATE-EARLY PROTEIN LIKELY TO PLAY A ROLE IN CELL
CC       GROWTH REGULATION (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: INCREASED EXPRESSION IN WILMS TUMOR OF THE
CC       STROMAL TYPE.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; X78351; G825696; -.
DR   EMBL; X78352; G825696; JOINED.
DR   EMBL; X78353; G825696; JOINED.
DR   EMBL; X78354; G825696; JOINED.
DR   EMBL; X96584; E228691; -.
DR   MIM; 164958; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   PROTO-ONCOGENE; GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL       1     27       POTENTIAL.
FT   CHAIN       28    357       NOV PROTEIN HOMOLOG.
FT   DOMAIN     108    174       VWFC.
FT   DOMAIN     264    338       CTCK.
FT   DISULFID   264    301       BY SIMILARITY.
FT   DISULFID   281    315       BY SIMILARITY.
FT   DISULFID   292    331       BY SIMILARITY.
FT   DISULFID   295    333       BY SIMILARITY.
FT   DISULFID   300    337       BY SIMILARITY.
FT   CARBOHYD    97     97       POTENTIAL.
FT   CARBOHYD   280    280       POTENTIAL.
SQ   SEQUENCE   357 AA;  39162 MW;  DA8B009D CRC32;
```

```
  Query Match             13.5%;  Score 123.5;  DB 1;  Length 357;
  Best Local Similarity   31.8%;  Pred. No. 2.9e-06;
  Matches   42;  Conservative   25;  Mismatches   34;  Indels   31;  Gaps    9;

Qy       7 LTTLLVPAHLVA--AWSNNYAVDCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGR 61
           || ||:  ||::  | :   : :|| :| :    ::| |    |:||| |:|| |||  |
Db      18 LTFLLL--HLLGQVAATQRCPPQCPGRCPA----TPPTCAPGVRAVLDGCSCCLVCARQR 71

Qy      62 GETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICK-----DCPYG--------TFG 108
           ||:|    | ::    ::|| |: |   || ::: |||      :| :          |
Db      72 GESC----SDLEPCDESSGLYCDRS--ADP-SNQTGICTAVEGDNCVFDGVIYRSGEKFQ 124

Qy     109 MDCRETCNCQSG 120
             |:  | |: |
Db     125 PSCKFQCTCRDG 136
```

```
RESULT   5
IBP4_BOVIN
ID   IBP4_BOVIN     STANDARD;      PRT;   258 AA.
AC   Q05716;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 4 PRECURSOR (IGFBP-4)
DE   (IBP-4) (IGF-BINDING PROTEIN 4).
GN   IGFBP4.
OS   BOS TAURUS (BOVINE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; ARTIODACTYLA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93125553.
RA   MOSER D.R., LOWE W.L. JR., DAKE B.L., BOOTH B.A., BOES M.,
RA   CLEMMONS D.R., BAR R.S.;
RL   MOL. ENDOCRINOL. 6:1805-1814(1992).
CC   -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS
CC       AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC       PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE
CC       INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.
CC   -!- BINDS IGF-II MORE THAN IGF-I.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY.
DR   EMBL; S52770; G263304; -.
DR   PIR; A45403; A45403.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW   GROWTH FACTOR BINDING; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL       1     21       BY SIMILARITY.
FT   CHAIN       22    258       INSULIN-LIKE GROWTH FACTOR BINDING
FT                               PROTEIN 4.
FT   CARBOHYD   125    125       POTENTIAL.
FT   DOMAIN     200    249       THYROGLOBULIN TYPE I.
SQ   SEQUENCE   258 AA;  27890 MW;  97880748 CRC32;
```

```
  Query Match             12.6%;  Score 115.5;  DB 1;  Length 258;
  Best Local Similarity   32.3%;  Pred. No. 1.6e-05;
  Matches   30;  Conservative   24;  Mismatches   30;  Indels    9;  Gaps    4;

Qy       1 MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDS--SECKSSPRCKRTVLD-DCGCCRVC 57
           | |: |:::||::|   :: ::: |: ||   :   : |:::  |   | :  ||||  |
Db       1 MLSLCLMAALLLAAGPGPSLGDE-AIHCPPCSEEKLARCRPPVGCEELVREPGCGCCATC 59

Qy      58 AAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           | |:|  |     |:   :||:|||| |: | :
Db      60 ALGKGMPC-----GVYTPRCGSGLRCYPPRGVE 87
```

```
RESULT   6
IBP5_MOUSE
ID   IBP5_MOUSE     STANDARD;      PRT;   271 AA.
AC   Q07079;
DT   01-OCT-1994 (REL. 30, CREATED)
DT   01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 5 PRECURSOR (IGFBP-5)
DE   (IBP-5) (IGF-BINDING PROTEIN 5).
GN   IGFBP5 OR IGFBP-5.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=MYOBLASTS;
RX   MEDLINE; 94042976.
RA   JAMES P.L., JONES S.B., BUSBY W.H. JR., CLEMMONS D.R., ROTWEIN P.;
RL   J. BIOL. CHEM. 268:22305-22312(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SPLEEN;
RX   MEDLINE; 94307727.
RA   KOU K., JENKINS N.A., GILBERT D.J., COPELAND N.G., ROTWEIN P.;
RL   GENOMICS 20:412-418(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=KIDNEY;
RX   MEDLINE; 95121750.
RA   SCHULLER A.G.P., GROFFEN C., VAN NECK J.W., ZWARTHOFF E.C.,
RA   DROP S.L.S.;
RL   MOL. CELL. ENDOCRINOL. 104:57-66(1994).
CC   -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS
CC       AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC       PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE
CC       INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- TISSUE SPECIFICITY: MOST ABUNDANT IN KIDNEY, UTERUS AND
CC       GASTROCNEMIUS MUSCLE.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY.
DR   EMBL; L12447; G293384; -.
```

```
DR   EMBL; U02025; G437125; -.
DR   EMBL; U02023; G437125; JOINED.
DR   EMBL; U02027; G437125; JOINED.
DR   EMBL; U02024; G437125; JOINED.
DR   EMBL; X81583; G550385; -.
DR   MGD; MGI:96440; IGFBP5.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    271       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                 PROTEIN 5.
FT   DOMAIN      214    262       THYROGLOBULIN TYPE I.
FT   CONFLICT    112    112       MISSING (IN REF. 2).
SQ   SEQUENCE   271 AA;  30372 MW;  12DC64CA CRC32;
```

```
  Query Match             12.5%;  Score 114.5;  DB 1;  Length 271;
  Best Local Similarity   28.7%;  Pred. No. 2.1e-05;
  Matches   27;  Conservative   20;  Mismatches   26;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::  |||: :::           ||:: : |: : |  : |         ||||
Db       7 LLLLAAYAVPAQGLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 57

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      58 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 86
```

```
RESULT   7
CTGF_MOUSE
ID   CTGF_MOUSE     STANDARD;      PRT;   348 AA.
AC   P29268;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   CONNECTIVE TISSUE GROWTH FACTOR PRECURSOR (CTGF) (FISP-12 PROTEIN).
GN   CTGF OR FISP12 OR FISP-12.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91363290.
RA   RYSECK R.-P., MACDONALD-BRAVO H., MATTEI M.-G., BRAVO R.;
RL   CELL GROWTH DIFFER. 2:225-233(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91229699.
RA   BRUNNER A., CHINN J., NEUBAUER M.G., PURCHIO A.F.;
RL   DNA CELL BIOL. 10:293-300(1991).
CC   -!- TISSUE SPECIFICITY: TESTIS, SPLEEN, KIDNEY, LUNG, HEART, AND BRAIN
CC       (LOWEST LEVEL IN TESTIS AND HIGHEST IN LUNG).
CC   -!- INDUCTION: BY GROWTH FACTORS.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; M70641; G193314; -.
DR   EMBL; M70642; G193316; -.
DR   EMBL; M80263; G201946; -.
DR   PIR; A53228; A53228.
DR   MGD; MGI:95537; FISP12.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26    348       CONNECTIVE TISSUE GROWTH FACTOR.
FT   DOMAIN      100    166       VWFC.
FT   DOMAIN      255    329       CTCK.
FT   DISULFID    255    292       BY SIMILARITY.
FT   DISULFID    272    306       BY SIMILARITY.
FT   DISULFID    283    322       BY SIMILARITY.
FT   DISULFID    286    324       BY SIMILARITY.
FT   DISULFID    291    328       BY SIMILARITY.
FT   CONFLICT    161    161       K -> E (IN REF. 2).
SQ   SEQUENCE   348 AA;  37793 MW;  EAB92BE0 CRC32;
```

```
  Query Match             12.2%;  Score 112;  DB 1;  Length 348;
  Best Local Similarity   30.4%;  Pred. No. 4.8e-05;
  Matches   34;  Conservative   15;  Mismatches   35;  Indels   28;  Gaps    7;

Qy      27 DCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGP--GL 81
           ||: :|: :   ::|:|       ||| ||||||||   || |        :   | |  ||
Db      27 DCSAQCQCAA-EAAPHCPAGVSLVLDGCGCCRVCAKQLGELC------TERDPCDPHKGL 79

Qy      82 RCQPSNGEDPFGEEFGICKD-----CPYG--------TFGMDCRETCNCQSG 120
           |: ::   |     |:|       | :|        :|   |:  | |  |
Db      80 FCDFGS---PANRKIGVCTAKDGAPCVFGGSVYRSGESFQSSCKYQCTCLDG 128
```

```
RESULT   8
IBP5_RAT
ID   IBP5_RAT       STANDARD;      PRT;   271 AA.
AC   P24594;
DT   01-MAR-1992 (REL. 21, CREATED)
DT   01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 5 PRECURSOR (IGFBP-5)
DE   (IBP-5) (IGF-BINDING PROTEIN 5).
GN   IGFBP5 OR IGFBP-5.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; RODENTIA.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 20-53.
RC   TISSUE=OVARY;
RX   MEDLINE; 91244847.
RA   SHIMASAKI S., SHIMONAKA M., ZHANG H.-P., LING N.;
RL   J. BIOL. CHEM. 266:10646-10653(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE; 93176146.
RA   ZHU X., LING N., SHIMASAKI S.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 190:1045-1052(1993).
CC   -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS
CC       AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC       PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE
CC       INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.
CC   -!- TISSUE SPECIFICITY: MOSTLY IN KIDNEY.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY.
DR   EMBL; M62781; G204746; -.
DR   EMBL; L08275; E73333; -.
DR   PIR; A40403; A40403.
DR   PIR; JC1463; JC1463.
DR   PIR; F40403; F40403.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    271       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                 PROTEIN 5.
FT   DOMAIN      214    262       THYROGLOBULIN TYPE I.
SQ   SEQUENCE   271 AA;  30298 MW;  0AA79506 CRC32;
```

```
  Query Match             12.0%;  Score 110.5;  DB 1;  Length 271;
  Best Local Similarity   28.7%;  Pred. No. 5.8e-05;
  Matches   27;  Conservative   20;  Mismatches   26;  Indels   21;  Gaps    3;
```

```
Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::  |||: :::           ||:: : |: : |  : |         ||||
Db       7 LLLLAACAVPAQGLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 57

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      58 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 86


RESULT   9
CE10_CHICK
ID   CE10_CHICK     STANDARD;      PRT;   375 AA.
AC   P19336;
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   CEF-10 PROTEIN PRECURSOR.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
OC   GALLIFORMES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 89145206.
RA   SIMMONS D.L., LEVY D.B., YANNONI Y., ERIKSON R.L.;
RL   PROC. NATL. ACAD. SCI. U.S.A. 86:1178-1182(1989).
CC   -!- FUNCTION: PROBABLE SECRETED REGULATORY PROTEIN.
CC   -!- INDUCTION: BY V-SRC.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; J04496; G211436; -.
DR   PIR; A41428; A41428.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL       1     22
FT   CHAIN       23    375       CEF-10 PROTEIN.
FT   DOMAIN      98    164       VWFC.
FT   DOMAIN     281    355       CTCK.
FT   DISULFID   281    318       BY SIMILARITY.
FT   DISULFID   298    332       BY SIMILARITY.
FT   DISULFID   309    348       BY SIMILARITY.
FT   DISULFID   312    350       BY SIMILARITY.
FT   DISULFID   317    354       BY SIMILARITY.
SQ   SEQUENCE   375 AA;  40651 MW;  68B4BC92 CRC32;


  Query Match             12.1%;  Score 111;  DB 1;  Length 375;
  Best Local Similarity   29.1%;  Pred. No. 6.5e-05;
  Matches   39;  Conservative   19;  Mismatches   40;  Indels   36;  Gaps    9;

Qy       6 LLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRT---VLDDCGCCRVCAAGRG 62
           | ::||  |:|  : :      ||  |: :   ::|:|      | | ||||:|||
Db      10 LAAALLCLARLALGSP------CPAVCQCPA--AAPQCAPGVGLVPDGCGCCKVCAKQLN 61

Qy      63 ETCYRTVSGMDGMKCGP--GLRCQPSNGEDPFGEEFGICKD------CPYG--------T 106
           | | || :      |    || |  : |  | : : |||:       | |          :
Db      62 EDCSRTQP------CDHTKGLEC--NFGASPAATN-GICRAQSEGRPCEYNSKIYQNGES 112

Qy     107 FGMDCRETCNCQSG 120
           |  :|:  | |  |
Db     113 FQPNCKHQCTCIDG 126


RESULT  10
IBP5_HUMAN
ID   IBP5_HUMAN      STANDARD;      PRT;   272 AA.
AC   P24593;
DT   01-MAR-1992 (REL. 21, CREATED)
DT   01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 5 PRECURSOR (IGFBP-5)
DE   (IBP-5) (IGF-BINDING PROTEIN 5).
GN   IGFBP5 OR IBP5.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=OSTEOSARCOMA;
RX   MEDLINE; 91207396.
RA   KIEFER M.C., IOH R.S., BAUER D.M., ZAPF J.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 176:219-225(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 91244847.
RA   SHIMASAKI S., SHIMONAKA M., ZHANG H.-P., LING N.;
RL   J. BIOL. CHEM. 266:10646-10653(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94193798.
RA   ALLANDER S.V., LARSSON C., EHRENBORG E., SUWANICHKUL A., WEBER G.,
RA   MORRIS S.L., BAJALICA S., KIEFER M.C., LUTHMAN H., POWELL D.R.;
RL   J. BIOL. CHEM. 269:10891-10898(1994).
RN   [4]
RP   SEQUENCE OF 24-43.
RX   MEDLINE; 91207395.
RA   ANDRESS D.L., BIRNBAUM R.S.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 176:213-218(1991).
CC   -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS
CC       AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC       PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE
CC       INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.
CC   -!- TISSUE SPECIFICITY: OSTEOSARCOMA, AND AT LOWER LEVELS IN LIVER,
CC       KIDNEY, AND BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY.
DR   EMBL; M65062; G184820; -.
DR   EMBL; M62782; G184818; -.
DR   EMBL; L27559; G505589; -.
DR   EMBL; L27556; G505589; JOINED.
DR   EMBL; L27557; G505589; JOINED.
DR   EMBL; L27558; G505589; JOINED.
DR   PIR; JH0391; JH0391.
DR   PIR; B40403; B40403.
DR   PIR; PH0143; PH0143.
DR   MIM; 146734; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL       1     20       POTENTIAL.
FT   CHAIN       21    272       INSULIN-LIKE GROWTH FACTOR BINDING
FT                               PROTEIN 5.
FT   DOMAIN     215    263       THYROGLOBULIN TYPE I.
SQ   SEQUENCE   272 AA;  30570 MW;  3A6AEEF2 CRC32;


  Query Match             11.9%;  Score 109.5;  DB 1;  Length 272;
  Best Local Similarity   27.7%;  Pred. No. 7.4e-05;
  Matches   26;  Conservative   20;  Mismatches   27;  Indels   21;  Gaps    3;

Qy       4 VLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLD-------DCGCCRV 56
           :|||::   ||: :::           ||:: : |: : |  : |         ||||
Db       8 LLLLAAYAGPAQSLGSFV---------HCEPCDEKALSMCPPSPLGCELVKEPGCGCCMT 58

Qy      57 CAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGED 90
           || : |::|     |:   :|: |||| |   |:
Db      59 CALAEGQSC-----GVYTERCAQGLRCLPRQDEE 87


RESULT  11
CYR6_HUMAN
```

```
ID   CYR6_HUMAN     STANDARD;      PRT;   381 AA.
AC   O00622; O14934;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   CYR61 PROTEIN PRECURSOR (GIG1 PROTEIN) (INSULIN-LIKE GROWTH FACTOR-
DE   BINDING PROTEIN 10).
GN   CYR61 OR GIG1 OR IGFBP10.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   ALBRECHT C., VON DER KAMMER H., KLAUDINY J., MAYHAUS M., NITSCH R.M.;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97280750.
RA   JAY P., BERGE-LEFRANC J.L., MARSOLLIER C., MEJEAN C., TAVIAUX S.,
RA   BERTA P.;
RL   ONCOGENE 14:1753-1757(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   KOLESNIKOVA T.V., LAU L.F.;
RL   SUBMITTED (JUN-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   BI A.B., YU L.;
RL   SUBMITTED (NOV-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: MAY ACT AS ONE OF THE MANY GROWTH FACTOR-BINDING
CC       PROTEINS; PROMOTES PROLIFERATION, MIGRATION AND ADHESION (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; Y12084; E311857; -.
DR   EMBL; U62015; G2130527; -.
DR   EMBL; AF003594; G2196782; -.
DR   EMBL; AF031385; G2606094; -.
DR   MIM; 602369; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   GROWTH FACTOR BINDING; SIGNAL.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25    381       CYR61 PROTEIN.
FT   DOMAIN       98    164       VWFC.
FT   DOMAIN      286    360       CTCK.
FT   DISULFID    286    323       BY SIMILARITY.
FT   DISULFID    303    337       BY SIMILARITY.
FT   DISULFID    314    353       BY SIMILARITY.
FT   DISULFID    317    355       BY SIMILARITY.
FT   DISULFID    322    359       BY SIMILARITY.
FT   CONFLICT    210    210       L -> I (IN REF. 4).
FT   CONFLICT    220    220       L -> R (IN REF. 4).
SQ   SEQUENCE   381 AA;  42026 MW;  2B091D9E CRC32;
```

```
  Query Match             12.0%;  Score 110.5;  DB 1;  Length 381;
  Best Local Similarity   26.1%;  Pred. No. 7.4e-05;
  Matches   36;  Conservative   21;  Mismatches   44;  Indels   37;  Gaps    8;

Qy       2 KSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRT---VLDDCGCCRVCA 58
           ::: |: |||  ::|  :        ||  |  :    :|:|      | | ||||:|||
Db       7 RALALVVTLLHLTRLALS-------TCPAACHCP--LEAPKCAPGVGLVRDGCGCCKVCA 57

Qy      59 AGRGETCYRTVSGMDGMKCGP--GLRCQPSNGEDPFGEEFGICKD------CPYG----- 106
             | | :| :      |    || |  : :  ::    |||:       | |
Db      58 KQLNEDCSKTQP------CDHTKGLECNFGASSTALK---GICRAQSEGRPCEYNSRIYQ 108

Qy     106 ---TFGMDCRETCNCQSG 120
               :|  :|:  | |  |
Db     109 NGESFQPNCKHQCTCIDG 126
```

```
RESULT  12
CTGF_HUMAN
ID   CTGF_HUMAN     STANDARD;      PRT;   349 AA.
AC   P29279;
DT   01-DEC-1992 (REL. 24, CREATED)
DT   01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   CONNECTIVE TISSUE GROWTH FACTOR PRECURSOR.
GN   CTGF.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=UMBILICAL VEIN ENDOTHELIAL CELLS;
RX   MEDLINE; 91373462.
RA   BRADHAM D.M., IGARASHI A., POTTER R.L., GROTENDORST G.R.;
RL   J. CELL BIOL. 114:1285-1294(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=UMBILICAL VEIN ENDOTHELIAL CELLS;
RX   MEDLINE; 93187114.
RA   IGARASHI A., BRADHAM D.M., OKOCHI H., GROTENDORST G.R.;
RL   J. DERMATOL. 19:642-643(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   OEMAR B.S., WERNER A., YANG Z., GARNIER J.M., GENTZ R., LUESCHER T.F.;
RL   SUBMITTED (APR-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: MAJOR CONNECTIVE TISSUE MITOATTRACTANT SECRETED BY
CC       HUMAN VASCULAR ENDOTHELIAL CELLS. THIS IMMEDIATE-EARLY PROTEIN
CC       MAY BIND ONE OF THE PDGF CELL SURFACE RECEPTORS.
CC   -!- SUBUNIT: MONOMER.
CC   -!- ALTERNATIVE PRODUCTS: A SHORTER FORM MAY BE PRODUCED BY
CC       ALTERNATIVE SPLICING OF THE SAME GENE.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY. CEF-10/CYR61/CTFG/FISP-12/NOV PROTEIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
DR   EMBL; M92934; G180924; -.
DR   EMBL; X78947; G474934; -.
DR   PIR; A40551; A40551.
DR   PIR; S44205; S44205.
DR   MIM; 121009; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
DR   PROSITE; PS01208; VWFC; 1.
KW   GROWTH FACTOR BINDING; SIGNAL; ALTERNATIVE SPLICING.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27    349       CONNECTIVE TISSUE GROWTH FACTOR.
FT   DOMAIN      101    167       VWFC.
FT   DOMAIN      256    330       CTCK.
FT   DISULFID    256    293       BY SIMILARITY.
FT   DISULFID    273    307       BY SIMILARITY.
FT   DISULFID    284    323       BY SIMILARITY.
FT   DISULFID    287    325       BY SIMILARITY.
FT   DISULFID    292    329       BY SIMILARITY.
FT   CARBOHYD     28     28       POTENTIAL.
FT   CARBOHYD    225    225       POTENTIAL.
FT   VARSPLIC    172    198       MISSING (IN SHORT FORM).
SQ   SEQUENCE   349 AA;  38069 MW;  C21E9662 CRC32;
```

```
  Query Match             11.9%;  Score 109;  DB 1;  Length 349;
  Best Local Similarity   30.4%;  Pred. No. 0.0001;
  Matches   34;  Conservative   14;  Mismatches   36;  Indels   28;  Gaps    7;

Qy      27 DCPQHCDSSECKSSPRCK---RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGP--GL 81
```

```
          :|:  |  ::  ::|||      ||| ||||||||  || |       :   | |  ||
Db      28 NCSGPCRCPD-EPAPRCPAGVSLVLDGCGCCRVCAKQLGELC------TERDPCDPHKGL 80

Qy      82 RCQPSNGEDPFGEEFGICKD-----CPYG--------TFGMDCRETCNCQSG 120
           |: ::   |     |:|       | :|        :|   |:  | |  |
Db      81 FCDFGS---PANRKIGVCTAKDGAPCIFGGTVYRSGESFQSSCKYQCTCLDG 129


RESULT  13
IBP4_SHEEP
ID   IBP4_SHEEP     STANDARD;      PRT;   237 AA.
AC   Q28893;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 4 (IGFBP-4) (IBP-4)
DE   (IGF-BINDING PROTEIN 4).
GN   IGFBP4.
OS   OVIS ARIES (SHEEP).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; ARTIODACTYLA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RX   MEDLINE; 95151165.
RA   CARR J.M., GRANT P.A., FRANCIS G.L., OWENS J.A., WALLACE J.C.,
RA   WALTON P.E.;
RL   J. MOL. ENDOCRINOL. 13:219-236(1994).
CC   -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS
CC       AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC       PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE
CC       INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY.
DR   EMBL; S77394; G944952; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW   GROWTH FACTOR BINDING; GLYCOPROTEIN.
FT   DOMAIN      179    228       THYROGLOBULIN TYPE I.
FT   CARBOHYD    104    104       POTENTIAL.
SQ   SEQUENCE   237 AA;  25869 MW;  C1C79FEA CRC32;


  Query Match             11.7%;  Score 107;  DB 1;  Length 237;
  Best Local Similarity   34.8%;  Pred. No. 0.00012;
  Matches   24;  Conservative   13;  Mismatches   24;  Indels    8;  Gaps    3;

Qy      25 AVDCPQHCDS--SECKSSPRCKRTVLD-DCGCCRVCAAGRGETCYRTVSGMDGMKCGPGL 81
           |: ||    :   : |:::  |   | :  ||||  || |:|  |     |:    ||:||
Db       3 AIHCPPCSEEKLARCRPPVGCEELVREPGCGCCATCALGKGMPC-----GVYTPDCGSGL 57

Qy      82 RCQPSNGED 90
           ||:|: | :
Db      58 RCHPPRGVE 66


RESULT  14
IBP4_HUMAN
ID   IBP4_HUMAN     STANDARD;      PRT;   258 AA.
AC   P22692;
DT   01-AUG-1991 (REL. 19, CREATED)
DT   01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 4 PRECURSOR (IGFBP-4)
DE   (IBP-4) (IGF-BINDING PROTEIN 4).
GN   IGFBP4 OR IBP4.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; PRIMATES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91186988.
RA   LATOUR D., MOHAN S., LINKHART T.A., BAYLINK D.J., STRONG D.D.;
RL   MOL. ENDOCRINOL. 4:1806-1814(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE; 91133415.
RA   SHIMASAKI S., UCHIYAMA F., SHIMONAKA M., LING N.;
RL   MOL. ENDOCRINOL. 4:1451-1458(1990).
RN   [3]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 22-41.
RC   TISSUE=OSTEOSARCOMA;
RX   MEDLINE; 91225006.
RA   KIEFER M.C., MASIARZ F.R., BAUER D.M., ZAPF J.;
RL   J. BIOL. CHEM. 266:9043-9049(1991).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RA   STRONG D.D., MORALES S., LEE K., BOONYARATANAKORNKIT V.,
RA   BAYLINK D.J., MOHAN S.;
RL   SUBMITTED (FEB-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN   [5]
RP   SEQUENCE OF 22-53.
RC   TISSUE=COLON;
RX   MEDLINE; 91235178.
RA   CULOUSCOU J.-M., SHOYAB M.;
RL   CANCER RES. 51:2813-2819(1991).
CC   -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS
CC       AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC       PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE
CC       INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.
CC   -!- BINDS IGF-II MORE THAN IGF-I.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY.
DR   EMBL; M38177; -; NOT_ANNOTATED_CDS.
DR   EMBL; M62403; G184816; -.
DR   EMBL; U20982; G695254; -.
DR   PIR; A36549; A36549.
DR   PIR; B37252; B37252.
DR   PIR; B39842; B39842.
DR   MIM; 146733; -.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW   GROWTH FACTOR BINDING; SIGNAL; GLYCOPROTEIN.
FT   SIGNAL        1     21
FT   CHAIN        22    258       INSULIN-LIKE GROWTH FACTOR BINDING
FT                                 PROTEIN 4.
FT   CARBOHYD    125    125       POTENTIAL.
FT   DOMAIN      200    249       THYROGLOBULIN TYPE I.
FT   CONFLICT     51     51       P -> A (IN REF. 1, 4 AND 5).
FT   CONFLICT    198    198       I -> F (IN REF. 1 AND 4).
SQ   SEQUENCE   258 AA;  27934 MW;  58AC8AC3 CRC32;


  Query Match             11.6%;  Score 106.5;  DB 1;  Length 258;
  Best Local Similarity   31.8%;  Pred. No. 0.00015;
  Matches   28;  Conservative   22;  Mismatches   29;  Indels    9;  Gaps    4;

Qy       6 LLTTLLVPAHLVAAWSNNYAVDCPQHCDS--SECKSSPRCKRTVLD-DCGCCRVCAAGRG 62
           |:::||::|   :: ::: |: ||    :   : |:::  |   | :  ||||  || | |
Db       6 LVAALLLAAGPGPSLGDE-AIHCPPCSEEKLARCRPPVGCEELVREPGCGCCATCALGLG 64

Qy      63 ETCYRTVSGMDGMKCGPGLRCQPSNGED 90
             |     |:   :||:|||| |: | :
Db      65 MPC-----GVYTPRCGSGLRCYPPRGVE 87


RESULT  15
IBP3_PIG
ID   IBP3_PIG       STANDARD;      PRT;   266 AA.
AC   P16611;
DT   01-AUG-1990 (REL. 15, CREATED)
DT   01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT   01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE   INSULIN-LIKE GROWTH FACTOR BINDING PROTEIN 3 (IGFBP-3) (IBP-3) (IGF-
```

```
DE   BINDING PROTEIN 3).
GN   IGFBP3.
OS   SUS SCROFA (PIG).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC   EUTHERIA; ARTIODACTYLA.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 90130475.
RA   SHIMASAKI S., SHIMONAKA M., UI M., INOUYE S., SHIBATA F., LING N.;
RL   J. BIOL. CHEM. 265:2198-2202(1990).
RN   [2]
RP   SEQUENCE OF 1-15.
RX   MEDLINE; 92109718.
RA   COLEMAN M.E., PAN Y.-C.E., ETHERTON T.D.;
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 181:1131-1136(1991).
CC   -!- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFS
CC       AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC       PROMOTING EFFECTS OF THE IGFS ON CELL CULTURE. THEY ALTER THE
CC       INTERACTION OF IGFS WITH THEIR CELL SURFACE RECEPTORS.
CC   -!- SUBUNIT: FORMS A TERNARY COMPLEX OF ABOUT 140 TO 150 KD WITH IGF-I
CC       OR IGF-II AND A 85 KD GLYCOPROTEIN (ALS).
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC   -!- BINDS IGF-II MORE THAN IGF-I.
CC   -!- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC       PROTEIN FAMILY.
DR   EMBL; J05228; G164501; -.
DR   PIR; A35037; A35037.
DR   PIR; JH0516; JH0516.
DR   HSSP; P17494; 1KST.
DR   PROSITE; PS00222; IGF_BINDING; 1.
DR   PROSITE; PS00484; THYROGLOBULIN_1; 1.
KW   GROWTH FACTOR BINDING; GLYCOPROTEIN.
FT   CARBOHYD     91     91       POTENTIAL.
FT   CARBOHYD    111    111       POTENTIAL.
FT   CARBOHYD    174    174       POTENTIAL.
FT   DOMAIN      211    260       THYROGLOBULIN TYPE I.
FT   CONFLICT      5      6       VG -> A (IN REF. 2).
SQ   SEQUENCE   266 AA;  28910 MW;  53E9830D CRC32;


  Query Match             11.4%;  Score 104.5;  DB 1;  Length 266;
  Best Local Similarity   31.3%;  Pred. No. 0.00025;
  Matches   26;  Conservative   18;  Mismatches   20;  Indels   19;  Gaps    4;

Qy      31 HCDSSECK---------SSPRCKRTVLD-DCGCCRVCAAGRGETCYRTVSGMDGMKCGPG 80
           :|:: : :          ::| |   | :  ||||  ||   |::|     |:   :||:|
Db      12 RCEPCDARALAQCAPPPAAPPCAELVREPGCGCCLTCALREGQAC-----GVYTERCGAG 66

Qy      81 LRCQPSNGE----DPFGEEFGIC 99
           |||||: ||    ::: :  |||
Db      67 LRCQPPPGEPRPLQALLDGRGIC 89


Search completed: May  4, 1999, 16:50:02
Job time: 10240 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 4, 1999, 02:47:21 ; Search time 36.71 Seconds
(without alignments)
276.522 Million cell updates/sec

Title: US-09-037-460-2
Perfect score: 184
Sequence: 1 MKSVLLLTTLLVPAHLVAAW.....EVVKENAAGSPVMRKWLNPR 184

Scoring table: OLIGO

Searched: 180763 seqs, 55169189 residues

Word = 30

Database :
SPTREMBL_8:*
1: sp_fungi:*
2: sp_human:*
3: sp_invertebrate:*
4: sp_mammal:*
5: sp_mhc:*
6: sp_organelle:*
7: sp_phage:*
8: sp_plant:*
9: sp_bacteria:*
10: sp_rodent:*
11: sp_virus:*
12: sp_vertebrate:*
13: sp_unclassified:*
14: sp_archea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	184	100.0	184	2	Q15330	Q15330 homo sapien

ALIGNMENTS

RESULT 1
Q15330
ID Q15330 PRELIMINARY; PRT; 184 AA.
AC Q15330;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ESM-1 SECRETORY PROTEIN PRECURSOR.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96355375.
RA LASSALLE P.M., MOLET S., JANIN A., VANDER-HEYDEN J.E., TAVERNIER J.,
RA FIERS W., DEVOS R.E., TONNEL A.E.B.;
RT "ESM-1 is a novel human endothelial cell-specific molecule expressed
RT in lung and regulated by cytokines.";
RL J. BIOL. CHEM. 271:20458-20464(1996).
DR EMBL; X89426; E189266; -.
DR PFAM; PF00219; IGFBP; 1.
KW SIGNAL.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 184 ESM-1 SECRETORY PROTEIN.
SQ SEQUENCE 184 AA; 20095 MW; 08D109DF CRC32;

Query Match 100.0%; Score 184; DB 2; Length 184;
Best Local Similarity 100.0%; Pred. No. 1.5e-180;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLDDCGCCRVCAAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLDDCGCCRVCAAG 60

Qy     61 RGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICKDCPYGTFGMDCRETCNCQSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 RGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICKDCPYGTFGMDCRETCNCQSG 120

Qy    121 ICDRGTGKCLKFPFFQYSVTKSSNRFVSLTEHDMASGDGNIVREEVVKENAAGSPVMRKW 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ICDRGTGKCLKFPFFQYSVTKSSNRFVSLTEHDMASGDGNIVREEVVKENAAGSPVMRKW 180

Qy    181 LNPR 184
          ||||
Db    181 LNPR 184
```

Search completed: May 4, 1999, 05:33:53
Job time: 9992 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 4, 1999, 09:00:24 ; Search time 26.81 Seconds
(without alignments)
138.809 Million cell updates/sec

Title: US-09-037-460-2
Perfect score: 184
Sequence: 1 MKSVLLLTTLLVPAHLVAAW.....EVVKENAAGSPVMRKWLNPR 184

Scoring table: OLIGO

Searched: 162890 seqs, 20225328 residues word = 30

Database : A_Geneseq_34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	184	100.0	184	1	R98994	Vascular IBP-like

ALIGNMENTS

RESULT 1
R98994
ID R98994 standard; Protein; 184 AA.
AC R98994;
DT 06-NOV-1996 (first entry)
DE Vascular IBP-like growth factor.
KW Vascular IBF-like growth factor; VIGF;
KW insulin-like growth factor binding protein; agonist; antagonist;
KW muscle wastage; osteoporosis; implant fixation; wound healing;
KW therapy; diagnosis.
OS Homo sapiens.
FH Key Location/Qualifiers
FT peptide 1. .21
FT /label= Sig_peptide
PN WO9617931-A1.
PD 13-JUN-1996.
PF 09-DEC-1994; U14388.
PR 09-DEC-1994; WO-U14388.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Hastings GA, Rosen CA;
DR WPI; 96-287176/29.
DR N-PSDB; T34991.
PT Human vascular insulin-like growth factor binding protein-like
PT growth factor, and its nucleic acid sequence and (ant)agonists -
PT used, e.g. to treat muscle wasting diseases or aid implant fixation,
PT or limit excess connective tissue prodn. during wound healing.
PS Claim 14; Page 43-44; 61pp; English.
CC Human vascular insulin-like growth factor binding protein-like
CC growth factor (R98994), or VIGF, is a protein of primarily
CC vascular origin that is structurally related to the IBP and CCN
CC protein families. It can be expressed in e.g. E. coli, CHO or
CC insect host cells using a vector incorporating a cDNA clone
CC (T34991), or its derivative, obtd. from human umbilical
CC endothelial cells. It is useful therapeutically e.g. for
CC treating muscle wasting diseases or osteoporosis, or can be used
CC to detect diseases associated with under- or over-expression of VIGF,
CC or to screen for antagonists useful during wound healing.

SQ Sequence 184 AA;

Query Match 100.0%; Score 184; DB 1; Length 184;
Best Local Similarity 100.0%; Pred. No. 2.9e-185;
Matches 184; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLDDCGCCRVCAAG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCKRTVLDDCGCCRVCAAG 60

Qy     61 RGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICKDCPYGTFGMDCRETCNCQSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 RGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICKDCPYGTFGMDCRETCNCQSG 120

Qy    121 ICDRGTGKCLKFPFFQYSVTKSSNRFVSLTEHDMASGDGNIVREEVVKENAAGSPVMRKW 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ICDRGTGKCLKFPFFQYSVTKSSNRFVSLTEHDMASGDGNIVREEVVKENAAGSPVMRKW 180

Qy    181 LNPR 184
          ||||
Db    181 LNPR 184
```

Search completed: May 4, 1999, 10:00:11
Job time: 3587 sec

OM nucleic - nucleic search, using sw model

Run on: May 4, 1999, 02:45:59 ; Search time 823 Seconds
(without alignments)
5524.986 Million cell updates/sec

Title: US-09-037-460-1
Perfect score: 1271
Sequence: 1 CTGCTTCCCACCAGCAAAGA.....GTTNCCTNTNTCTNAAGGAC 1271

Scoring table: OLIGO_NUC

Searched: 808301 seqs, 1788773984 residues

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl1:*
9: gb_pl2:*
10: gb_pr1:*
11: gb_pr2:*
12: gb_pr3:*
13: gb_ro:*
14: gb_st:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: gb_htg:*
19: em_ba:*
20: em_fun:*
21: em_hum1:*
22: em_hum2:*
23: em_in:*
24: em_om:*
25: em_or:*
26: em_ov:*
27: em_pat:*
28: em_ph:*
29: em_pl:*
30: em_ro:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: em_htg:*
35: em_sts:*
36: gb_ba1:*
37: gb_ba2:*
38: gb_pl1:*
39: gb_pl2:*
40: gb_pr1:*
41: gb_pr2:*
42: gb_pr3:*
43: gb_sts:*

word size = 30

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

length
% of match

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1271	100.0	1271	6	AR004664	AR004664 Sequence
2	608	47.8	2006	10	HSRNAESM1	X89426 H.sapiens m
3	608	47.8	2006	40	HSRNAESM1	X89426 H.sapiens m

ALIGNMENTS

RESULT 1
AR004664
LOCUS AR004664 1271 bp DNA PAT 04-DEC-1998
DEFINITION Sequence 1 from patent US 5747280.
ACCESSION AR004664
NID g3965543
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1271)
AUTHORS Hastings,G.A. and Rosen,C.A.
TITLE Human vascular IBP-like growth factor
JOURNAL Patent: US 5747280-A 1 05-MAY-1998;
FEATURES Location/Qualifiers
source 1. .1271
/organism="unknown"
BASE COUNT 361 a 246 c 336 g 320 t 8 others
ORIGIN

Query Match 100.0%; Score 1271; DB 6; Length 1271;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CTGCTTCCCACCAGCAAAGACCACGACTGGAGAGCCGAGCCGGAGCAGCTGGGAAACATG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTGCTTCCCACCAGCAAAGACCACGACTGGAGAGCCGAGCCGGAGCAGCTGGGAAACATG 60

Qy      61 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCTGGTGGCCGCCTGGAGC 120

Qy     121 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGAGTGCAAAAGCAGCCCG 180

Qy     181 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGTGTGCGCTGCAGGGCGG 240

Qy     241 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAAGTGTGGCCCGGGGCTG 300

Qy     301 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTTTGGTATCTGCAAAGAC 360

Qy     361 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAACTGCCAGTCAGGCATC 420

Qy     421 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCAATATTCAGTAACCAAG 480

Qy     481 TCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATCTGGAGATGGCAATATT 540

Qy     541 GTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGTAATGAGGAAATGGTTA 600
```

```
Qy    601 AATCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAGGCTCTATTTTCGTGAYTGTTCAA 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AATCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAGGCTCTATTTTCGTGAYTGTTCAA 660

Qy    661 CACACAGCCAACATTTTAGGAACTTTCTAGATTATAGCATAAGGACATGTAATTTTTGAA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CACACAGCCAACATTTTAGGAACTTTCTAGATTATAGCATAAGGACATGTAATTTTTGAA 720

Qy    721 GACCAAATGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAATCCATAA 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GACCAAATGTGATGCATGGTGGATCCAGAAAACAAAAAGTAGGATACTTACAATCCATAA 780

Qy    781 CATCCATATGACTGAACACTTGTATGTGTTTGTTAAATATTCGAATGCATGTAGATTTGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CATCCATATGACTGAACACTTGTATGTGTTTGTTAAATATTCGAATGCATGTAGATTTGT 840

Qy    841 TAAATGTGTGTGTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTACATG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TAAATGTGTGTGTATAGTAACACTGAAGAACTAAAAATGCAATTTAGGTAATCTTACATG 900

Qy    901 GAGACAGGTCAACCAAAGAGGGAGCTAGGCAAAGCTGAAGACCGCAGTGAGTCAAATTAG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GAGACAGGTCAACCAAAGAGGGAGCTAGGCAAAGCTGAAGACCGCAGTGAGTCAAATTAG 960

Qy    961 TTCTTTGACTTTGATGTACATTAATGTTGGGATATGGAATGAAGACTTAAGAGCAGGAGA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TTCTTTGACTTTGATGTACATTAATGTTGGGATATGGAATGAAGACTTAAGAGCAGGAGA 1020

Qy   1021 AGATGGGGAGGGGGTGGGAGTGGGAAATAAAATATTTAGCCCTTCCTTGGTAGGTAGCTT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AGATGGGGAGGGGGTGGGAGTGGGAAATAAAATATTTAGCCCTTCCTTGGTAGGTAGCTT 1080

Qy   1081 CTCTAGAATTTAATTRTGCTTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAA 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CTCTAGAATTTAATTRTGCTTTTTTTTTTTTTTTTGGGCTTTGGGAAAAGTCAAAATAAA 1140

Qy   1141 ACAACCAGAAAACCCCTGAAGGAAGTAAGATGTTTGAAGCTTATGGAAATTTGAGTAACA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 ACAACCAGAAAACCCCTGAAGGAAGTAAGATGTTTGAAGCTTATGGAAATTTGAGTAACA 1200

Qy   1201 AACAGCTTTGANCTGAGAGCAATTYCAAAAGGCTGCTGATGTAGCCCCCGGGTTNCCTNT 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 AACAGCTTTGANCTGAGAGCAATTYCAAAAGGCTGCTGATGTAGCCCCCGGGTTNCCTNT 1260

Qy   1261 NTCTNAAGGAC 1271
          ||||||||||||
Db   1261 NTCTNAAGGAC 1271
```

```
RESULT   2
HSRNAESM1
LOCUS       HSRNAESM1    2006 bp    RNA            PRI       07-OCT-1996
DEFINITION  H.sapiens mRNA for ESM-1 protein.
ACCESSION   X89426
NID         g1150418
KEYWORDS    ESM-1 protein.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2006)
  AUTHORS   Lassalle,P., Molet,S., Janin,A., Heyden,J.V., Tavernier,J.,
            Fiers,W., Devos,R. and Tonnel,A.B.
  TITLE     ESM-1 is a novel human endothelial cell-specific molecule expressed
            in lung and regulated by cytokines
  JOURNAL   J. Biol. Chem. 271 (34), 20458-20464 (1996)
  MEDLINE   96355375
REFERENCE   2  (bases 1 to 2006)
  AUTHORS   Lassalle,P.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-1995) P.M. Lassalle, INSERM, Unite 416, 1, bd du
            Prof. CALMETTE, LILLE 59019, FRANCE
FEATURES             Location/Qualifiers
     source          1. .2006
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="HUVEC"
                     /clone="A11.1"
     sig_peptide     56. .112
     CDS             56. .610
                     /codon_start=1
                     /product="ESM-1 secretory protein"
                     /db_xref="PID:e189266"
                     /db_xref="PID:g1150419"
                     /translation="MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCK
                     RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICKDC
                     PYGTFGMDCRETCNCQSGICDRGTGKCLKFPFFQYSVTKSSNRFVSLTEHDMASGDGN
                     IVREEVVKENAAGSPVMRKWLNPR"
BASE COUNT      623 a    333 c    475 g    575 t
ORIGIN
```

```
  Query Match             47.8%;  Score 608;  DB 10;  Length 2006;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  608;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     45 GCAGCTGGGAAACATGAAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCT 104
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     43 GCAGCTGGGAAACATGAAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCT 102

Qy    105 GGTGGCCGCCTGGAGCAATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGA 164
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    103 GGTGGCCGCCTGGAGCAATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGA 162

Qy    165 GTGCAAAAGCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGT 224
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    163 GTGCAAAAGCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGT 222

Qy    225 GTGCGCTGCAGGGCGGGGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAA 284
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    223 GTGCGCTGCAGGGCGGGGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAA 282

Qy    285 GTGTGGCCCGGGGCTGAGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTT 344
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    283 GTGTGGCCCGGGGCTGAGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTT 342

Qy    345 TGGTATCTGCAAAGACTGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAA 404
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    343 TGGTATCTGCAAAGACTGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAA 402

Qy    405 CTGCCAGTCAGGCATCTGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCA 464
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    403 CTGCCAGTCAGGCATCTGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCA 462

Qy    465 ATATTCAGTAACCAAGTCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATC 524
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    463 ATATTCAGTAACCAAGTCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATC 522

Qy    525 TGGAGATGGCAATATTGTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGT 584
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    523 TGGAGATGGCAATATTGTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGT 582

Qy    585 AATGAGGAAATGGTTAAATCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAGGCTCTAT 644
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    583 AATGAGGAAATGGTTAAATCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAGGCTCTAT 642

Qy    645 TTTCGTGA 652
          ||||||||
Db    643 TTTCGTGA 650
```

```
RESULT   3
HSRNAESM1
LOCUS       HSRNAESM1    2006 bp    RNA            PRI       07-OCT-1996
DEFINITION  H.sapiens mRNA for ESM-1 protein.
ACCESSION   X89426
```

```
NID         g1150418
KEYWORDS    ESM-1 protein.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2006)
  AUTHORS   Lassalle,P., Molet,S., Janin,A., Heyden,J.V., Tavernier,J.,
            Fiers,W., Devos,R. and Tonnel,A.B.
  TITLE     ESM-1 is a novel human endothelial cell-specific molecule expressed
            in lung and regulated by cytokines
  JOURNAL   J. Biol. Chem. 271 (34), 20458-20464 (1996)
  MEDLINE   96355375
REFERENCE   2  (bases 1 to 2006)
  AUTHORS   Lassalle,P.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-1995) P.M. Lassalle, INSERM, Unite 416, 1, bd du
            Prof. CALMETTE, LILLE 59019, FRANCE
FEATURES             Location/Qualifiers
     source          1. .2006
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="HUVEC"
                     /clone="A11.1"
     sig_peptide     56. .112
     CDS             56. .610
                     /codon_start=1
                     /product="ESM-1 secretory protein"
                     /db_xref="PID:e189266"
                     /db_xref="PID:g1150419"
                     /translation="MKSVLLLTTLLVPAHLVAAWSNNYAVDCPQHCDSSECKSSPRCK
                     RTVLDDCGCCRVCAAGRGETCYRTVSGMDGMKCGPGLRCQPSNGEDPFGEEFGICKDC
                     PYGTFGMDCRETCNCQSGICDRGTGKCLKFPFFQYSVTKSSNRFVSLTEHDMASGDGN
                     IVREEVVKENAAGSPVMRKWLNPR"
BASE COUNT      623 a    333 c    475 g    575 t
ORIGIN
```

Query Match 47.8%; Score 608; DB 40; Length 2006;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 608; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      45 GCAGCTGGGAAACATGAAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCT 104
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      43 GCAGCTGGGAAACATGAAGAGCGTCTTGCTGCTGACCACGCTCCTCGTGCCTGCACACCT 102

Qy     105 GGTGGCCGCCTGGAGCAATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGA 164
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     103 GGTGGCCGCCTGGAGCAATAATTATGCGGTGGACTGCCCTCAACACTGTGACAGCAGTGA 162

Qy     165 GTGCAAAAGCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGT 224
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     163 GTGCAAAAGCAGCCCGCGCTGCAAGAGGACAGTGCTCGACGACTGTGGCTGCTGCCGAGT 222

Qy     225 GTGCGCTGCAGGGCGGGGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAA 284
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     223 GTGCGCTGCAGGGCGGGGAGAAACTTGCTACCGCACAGTCTCAGGCATGGATGGCATGAA 282

Qy     285 GTGTGGCCCGGGGCTGAGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTT 344
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     283 GTGTGGCCCGGGGCTGAGGTGTCAGCCTTCTAATGGGGAGGATCCTTTTGGTGAAGAGTT 342

Qy     345 TGGTATCTGCAAAGACTGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAA 404
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     343 TGGTATCTGCAAAGACTGTCCCTACGGCACCTTCGGGATGGATTGCAGAGAGACCTGCAA 402

Qy     405 CTGCCAGTCAGGCATCTGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCA 464
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     403 CTGCCAGTCAGGCATCTGTGACAGGGGGACGGGAAAATGCCTGAAATTCCCCTTCTTCCA 462

Qy     465 ATATTCAGTAACCAAGTCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATC 524
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     463 ATATTCAGTAACCAAGTCTTCCAACAGATTTGTTTCTCTCACGGAGCATGACATGGCATC 522

Qy     525 TGGAGATGGCAATATTGTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGT 584
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     523 TGGAGATGGCAATATTGTGAGAGAAGAAGTTGTGAAAGAGAATGCTGCCGGGTCTCCCGT 582

Qy     585 AATGAGGAAATGGTTAAATCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAGGCTCTAT 644
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     583 AATGAGGAAATGGTTAAATCCACGCTGATCCCGGCTGTGATTTCTGAGAGAAGGCTCTAT 642

Qy     645 TTTCGTGA 652
           ||||||||
Db     643 TTTCGTGA 650
```

Search completed: May 4, 1999, 04:20:37
Job time: 5678 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: May 4, 1999, 08:18:16 ; Search time 25.07 Seconds
(without alignments)
22.413 Million cell updates/sec

Title: US-09-037-460-2_COPY_55_69
Perfect score: 78
Sequence: 1 RVCAAGRGETCYRTV 15

Scoring table: PAM150

Searched: 116738 seqs, 37460341 residues

Database : PIR_58:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	40	51.3	115	2	I37901	Histone H2A relate
2	42	53.8	607	1	ABHOS	serum albumin prec
3	39	50.0	218	2	A70080	two-component resp
4	36	46.2	107	1	A60361	neuroparsin A prec
5	36	46.2	136	2	S78428	destabilase 2 - me
6	31	39.7	25	2	JH0701	omega-conotoxin MV
7	29	37.2	12	2	S26553	T-cell receptor be
8	33.5	42.9	81	2	JC4621	cardiotoxin N prec
9	38	48.7	599	2	A26100	genome polyprotein
10	29.5	37.8	17	2	A38824	tachyplesin I - ho
11	29.5	37.8	17	2	JX0125	tachyplesin III -
12	29.5	37.8	17	2	A30068	tachyplesin - hors
13	29.5	37.8	17	2	JU0123	tachyplesin II - h
14	35	44.9	185	2	S59504	ferric pseudobacti
15	29.5	37.8	19	2	JX0124	tachyplesin I prec
16	39	50.0	1040	2	E71412	hypothetical prote
17	34	43.6	130	2	S52116	transforming prote
18	38	48.7	728	1	JH0579	hepatocyte growth
19	33.5	42.9	122	2	A25806	phospholipase A2 (
20	36	46.2	351	2	S20078	NOV protein - chic
21	36	46.2	357	2	I38069	gene novH protein
22	28	35.9	13	2	B58533	CD61 homolog - cha
23	36	46.2	387	3	JC5949	galanin receptor 2
24	36	46.2	388	2	C70605	probable aminotran
25	33.5	42.9	138	1	PSVIAA	phospholipase A2 (
26	33.5	42.9	138	1	PSVIAC	phospholipase A2 (
27	33.5	42.9	138	2	S10333	ammodytoxin B prec
28	33.5	42.9	138	2	I51386	phospholipase A2 (
29	36	46.2	395	2	S40685	probable G protein
30	31.5	40.4	60	1	H3NJ2I	cytotoxin 2 - Naja
31	31.5	40.4	60	1	H3NJ2R	cytotoxin 2 - Asia
32	31.5	40.4	60	1	H3NJ2F	cytotoxin 2 - Chin
33	31.5	40.4	60	1	H3NJ4F	cytotoxin 4 - Chin
34	31.5	40.4	60	1	H3NJ5F	cytotoxin 5 - Chin
35	31.5	40.4	60	1	H3NJ2K	cytotoxin 2 - mono
36	31.5	40.4	60	1	H3NJ3K	cytotoxin 3 - mono
37	31.5	40.4	60	1	H3NJ1I	cytotoxin 1 - Naja
38	31.5	40.4	60	1	H3NJ1K	cytotoxin 1 - mono
39	31.5	40.4	60	2	JS0028	cytotoxin II - mon
40	31.5	40.4	60	2	JK0222	cytotoxin 10 - mon
41	36	46.2	457	2	G71124	hypothetical prote
42	35	44.9	301	2	S63620	cymI protein - Kle
43	36	46.2	472	1	A35327	vitamin D-binding
44	37	47.4	728	1	A35644	hepatocyte growth
45	37	47.4	728	1	A60185	hepatocyte growth

ALIGNMENTS

RESULT 1
I37901
Histone H2A related - human
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Aug-1998
C;Accession: I37901
R;Naylor, J.A.; Buck, D.; Green, P.M.; Williamson, H.; Bentley, D.; Giannelli, F.
Hum. Mol. Genet. 4, 1217-1224, 1995
A;Title: Investigation of the Factor VIII intron 22 repeated region (int22h) and the
A;Reference number: I37901
A;Accession: I37901
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-115 <RES>
A;Cross-references: EMBL:X86012; NID:g976082; PID:g976083
C;Superfamily: histone H2A

Query Match 51.3%; Score 40; DB 2; Length 115;
Best Local Similarity 75.0%; Pred. No. 3.7;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy       4 AAGRGETCYRTV 15
           |:||| || |||
Db      13 AGGRGRTCSRTV 24
```

RESULT 2
ABHOS
serum albumin precursor - horse
C;Species: Equus caballus (domestic horse)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Sep-1997
C;Accession: S34053
R;Ho, J.X.; Holowachuk, E.W.; Norton, E.J.; Twigg, P.D.; Carter, D.C.
Eur. J. Biochem. 215, 205-212, 1993
A;Title: X-ray and primary structure of horse serum albumin (Equus caballus) at 0.27-
A;Reference number: S34053
A;Accession: S34053
A;Molecule type: mRNA
A;Residues: 1-607 <HOA>
A;Cross-references: GB:X74045; NID:g399671; PID:g399672
C;Comment: Serum albumin is synthesized in the liver as preproalbumin. It binds coppe
teroid hormones (weak bonds with these hormones promote their transfer across the mem
C;Superfamily: serum albumin; serum albumin repeat homology
C;Keywords: carrier protein; duplication; metal binding; plasma
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-24/Domain: propeptide #status predicted <PRO>
F;25-607/Product: serum albumin #status predicted <MAT>
F;29-201/Domain: serum albumin repeat homology <SA1>
F;220-393/Domain: serum albumin repeat homology <SA2>
F;412-591/Domain: serum albumin repeat homology <SA3>
F;27/Binding site: copper (His) #status predicted
F;77-86,99-115,114-125,147-192,191-200,223-269,268-276,288-302,301-312,339-384,383-39
F;263/Binding site: bilirubin (Lys) #status predicted

Query Match 53.8%; Score 42; DB 1; Length 607;
Best Local Similarity 46.7%; Pred. No. 8;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

```
Qy       1 RVCAAGRGETCYRTV 15
           : || :   :|||||
```

```
Db    382 KCCAEADPPACYRTV 396


RESULT   3
A70080
two-component response regulator [YxjM] homolog yxjL - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998
C;Accession: A70080
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.;
Nature 390, 249-256, 1997
A;Authors: Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
wood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.;
Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber,
A;Authors: Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigu
, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwol
 Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleic
A;Authors: Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror
amakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol
.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa,
A;Authors: Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033
A;Accession: A70080
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-218 <KUN>
A;Cross-references: GB:Z99123; GB:AL009126; NID:g2636240; PID:e1186390; PID:g2636426
A;Experimental source: strain 168
C;Genetics:
A;Gene: yxjL
C;Superfamily: regulatory protein comA; response regulator homology
F;8-119/Domain: response regulator homology <RRH>
F;58/Binding site: phosphate (Asp) (covalent) #status predicted


  Query Match            50.0%;  Score 39;  DB 2;  Length 218;
  Best Local Similarity  66.7%;  Pred. No. 9.9;
  Matches    8;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      4 AAGRGETCYRTV 15
          ||:|||: :|||
Db    121 AAARGEAIFRTV 132


RESULT   4
A60361
neuroparsin A precursor - migratory locust
N;Contains: neuroparsin A; neuroparsin B
C;Species: Locusta migratoria (migratory locust)
C;Date: 30-Jun-1993 #sequence_revision 12-Apr-1996 #text_change 12-Apr-1996
C;Accession: A61618; A60361; S03545
R;Lagueux, M.; Kromer, E.; Girardie, J.
Insect Biochem. Mol. Biol. 22, 511-516, 1992
A;Title: Cloning of a Locusta cDNA encoding neuroparsin A.
A;Reference number: A61618
A;Accession: A61618
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-107 <LAG>
R;Girardie, J.; Huet, J.C.; Pernollet, J.C.
Insect Biochem. 20, 659-666, 1990
A;Title: The locust neuroparsin A: sequence and similarities with vertebrate and insect
A;Reference number: A60361
A;Accession: A60361
A;Molecule type: protein
A;Residues: 25-107 <GIR>
R;Girardie, J.; Girardie, A.; Huet, J.C.; Pernollet, J.C.
FEBS Lett. 245, 4-8, 1989
A;Title: Amino acid sequence of locust neuroparsins.
A;Reference number: S03545; MUID:89171328
A;Accession: S03545
A;Molecule type: protein
A;Residues: 30-107 <GI2>
C;Comment: Intermediate forms of the neuroparsin monomer are also found. It is not kn
C;Superfamily: neuroparsin
C;Keywords: disulfide bond; homodimer; hormone
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-24/Domain: propeptide #status experimental <PRO>
F;25-107/Product: neuroparsin A monomer #status experimental <MAT1>
F;30-107/Product: neuroparsin B monomer #status experimental <MAT2>


  Query Match            46.2%;  Score 36;  DB 1;  Length 107;
  Best Local Similarity  54.5%;  Pred. No. 17;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 RVCAAGRGETC 11
            :||| | |: |
Db     55 KVCAKGPGDKC 65


RESULT   5
S78428
destabilase 2 - medicinal leech
C;Species: Hirudo medicinalis (medicinal leech)
C;Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 13-Mar-1998
C;Accession: S78428
R;Fradkov, A.; Berezhnoy, S.; Barsova, E.V.; Zavalova, L.; Lukyanov, S.; Baskova, I.;
FEBS Lett. 390, 145-148, 1996
A;Title: Enzyme from the medicinal leech (Hirudo medicinalis) that specifically split
A;Reference number: S78427
A;Accession: S78428
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-136 <FRA>
A;Cross-references: EMBL:U24122; NID:g1255717; PID:g1255718


  Query Match            46.2%;  Score 36;  DB 2;  Length 136;
  Best Local Similarity  54.5%;  Pred. No. 21;
  Matches    6;  Conservative    2;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 RVCAAGRGETC 11
            | |::||  ||
Db     95 RFCTGGRTPTC 105


RESULT   6
JH0701
omega-conotoxin MVIIB - cone shell (Conus magus)
C;Species: Conus magus (magus cone)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 23-May-1997
C;Accession: JH0701; B34115
R;Hillyard, D.R.; Monje, V.D.; Mintz, I.M.; Bean, B.P.; Nadasdi, L.; Ramachandran, J.
Neuron 9, 69-77, 1992
A;Title: A new conus peptide ligand for mammalian presynaptic Ca2+ channels.
A;Reference number: JH0699; MUID:92337922
A;Accession: JH0701
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-25 <HIL>
R;Olivera, B.M.; Cruz, L.J.; de Santos, V.; LeCheminant, G.W.; Griffin, D.; Zeikus, R
Biochemistry 26, 2086-2090, 1987
A;Title: Neuronal calcium channel antagonists. Discrimination between calcium channel
A;Reference number: A34115; MUID:87299637
A;Accession: B34115
A;Molecule type: protein
A;Residues: 1-25 <OLI>
C;Superfamily: omega-conotoxin
C;Keywords: acetylcholine release inhibition; amidated carboxyl end; calcium channel
F;1-16,8-20,15-25/Disulfide bonds: #status predicted
F;25/Modified site: amidated carboxyl end (Cys) #status predicted
```

```
  Query Match             39.7%;  Score 31;  DB 2;  Length 25;
  Best Local Similarity   55.6%;  Pred. No. 30;
  Matches    5;  Conservative    2;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       6 GRGETCYRT 14
           |:| :| ||
Db       3 GKGASCHRT 11


RESULT  7
S26553
T-cell receptor beta chain (clone Cw3/56.1) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence_revision 17-Apr-1998 #text_change 17-Apr-1998
C;Accession: S26553
R;Casanova, J.L.; Cerottini, J.C.; Matthes, M.; Necker, A.; Gournier, H.; Barra, C.; Wid
J. Exp. Med. 176, 439-447, 1992
A;Title: H-2-restricted cytolytic T lymphocytes specific for HLA display T cell receptor
A;Reference number: S26512
A;Accession: S26553
A;Molecule type: mRNA
A;Residues: 1-12 <CAS>
A;Cross-references: EMBL:X68003
A;Experimental source: cytolytic T-lymphocyte, clone Cw3/56.1
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: T-cell receptor


  Query Match             37.2%;  Score 29;  DB 2;  Length 12;
  Best Local Similarity   60.0%;  Pred. No. 33;
  Matches    6;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       3 CAAGRGETCY 12
           ||:::||| |
Db       1 CASSWGETLY 10


RESULT  8
JC4621
cardiotoxin N precursor - Chinese cobra
N;Alternate names: ctxN
C;Species: Naja naja atra (Chinese cobra)
C;Date: 10-Apr-1996 #sequence_revision 24-May-1996 #text_change 25-Apr-1997
C;Accession: JC4621; JC2469
R;Chang, L.S.; Wu, P.F.; Lin, J.
Biochem. Biophys. Res. Commun. 219, 116-121, 1996
A;Title: cDNA sequence analysis and expression of cardiotoxins from Taiwan cobra.
A;Reference number: JC4619; MUID:96190679
A;Accession: JC4621
A;Molecule type: mRNA
A;Residues: 1-81 <CHA>
A;Cross-references: EMBL:Z54230; NID:g1054814; PID:g1000509
A;Experimental source: venom glands
R;Chiou, S.H.; Hung, C.C.; Huang, H.C.; Chen, S.T.; Wang, K.T.; Yang, C.C.
Biochem. Biophys. Res. Commun. 206, 22-32, 1994
A;Title: Sequence comparison and computer modelling of cardiotoxins and cobrotoxin isola
A;Reference number: JC2469
A;Accession: JC2469
A;Molecule type: protein
A;Residues: 22-81 <CHI>
A;Note: conformation by (1)H-NMR
C;Superfamily: snake toxin
C;Keywords: cardiotoxin; hemolysis; venom
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-81/Product: cardiotoxin N #status experimental <MAT>
F;24-42,35-59,63-74,75-80/Disulfide bonds: #status predicted


  Query Match             42.9%;  Score 33.5;  DB 2;  Length 81;
  Best Local Similarity   46.2%;  Pred. No. 34;
  Matches    6;  Conservative    4;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 RVCAAGRGETCYR 13
           : |||| : ||:
Db      33 KTCAAGK-NLCYK 44


RESULT  9
A26100
genome polyprotein - murine poliovirus (fragment)
N;Contains: proteinase (EC 3.4.-.-); RNA-directed RNA polymerase (EC 2.7.7.48)
C;Species: murine poliovirus, Theiler's encephalomyelitis virus
C;Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 05-Dec-1997
C;Accession: A26100
R;Ozden, S.; Tangy, F.; Chamorro, M.; Brahic, M.
J. Virol. 60, 1163-1165, 1986
A;Title: Theiler's virus genome is closely related to that of encephalomyocarditis vi
A;Reference number: A26100; MUID:87061197
A;Accession: A26100
A;Molecule type: genomic RNA
A;Residues: 1-599 <OZD>
A;Cross-references: GB:M14703; NID:g335241; PID:g335242
C;Superfamily: foot-and-mouth disease virus genome polyprotein
C;Keywords: hydrolase; nucleotidyltransferase; polyprotein; proteinase
F;1-139/Product: proteinase (fragment) #status predicted <PTS>
F;140-599/Product: RNA-directed RNA polymerase #status predicted <RPS>


  Query Match             48.7%;  Score 38;  DB 2;  Length 599;
  Best Local Similarity   55.6%;  Pred. No. 38;
  Matches    5;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5 AGRGETCYR 13
           ::| :||||
Db      38 PARNDTCYR 46


RESULT  10
A38824
tachyplesin I - horseshoe crab (Tachypleus gigas)
C;Species: Tachypleus gigas
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 11-Jul-1997
C;Accession: A38824
R;Muta, T.; Fujimoto, T.; Nakajima, H.; Iwanaga, S.
J. Biochem. 108, 261-266, 1990
A;Title: Tachyplesins isolated from hemocytes of southeast Asian horseshoe crabs (Car
ssing intermediate of its precursor.
A;Reference number: JX0124; MUID:91035357
A;Accession: A38824
A;Molecule type: protein
A;Residues: 1-17 <MUT>
A;Experimental source: hemocyte
C;Keywords: amidated carboxyl end
F;3-16,7-12/Disulfide bonds: #status predicted
F;17/Modified site: amidated carboxyl end (Arg) #status experimental


  Query Match             37.8%;  Score 29.5;  DB 2;  Length 17;
  Best Local Similarity   53.8%;  Pred. No. 38;
  Matches    7;  Conservative    0;  Mismatches    3;  Indels    3;  Gaps    1;

Qy       1 RVCAAGRGETCYR 13
           |||  |    |||
Db       5 RVCYRG---ICYR 14


RESULT  11
JX0125
tachyplesin III - horseshoe crab (Tachypleus gigas)
C;Species: Tachypleus gigas
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 11-Jul-1997
C;Accession: JX0125
R;Muta, T.; Fujimoto, T.; Nakajima, H.; Iwanaga, S.
J. Biochem. 108, 261-266, 1990
A;Title: Tachyplesins isolated from hemocytes of southeast Asian horseshoe crabs (Car
```

ssing intermediate of its precursor.
A;Reference number: JX0124; MUID:91035357
A;Accession: JX0125
A;Molecule type: protein
A;Residues: 1-17 <MUT>
A;Experimental source: hemocyte
C;Keywords: amidated carboxyl end
F;3-16,7-12/Disulfide bonds: #status predicted
F;17/Modified site: amidated carboxyl end (Arg) #status experimental

Query Match 37.8%; Score 29.5; DB 2; Length 17;
Best Local Similarity 53.8%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

```
Qy       1 RVCAAGRGETCYR 13
           |||  |    |||
Db       5 RVCYRG---ICYR 14
```

RESULT 12
A30068
tachyplesin - horseshoe crab (Tachypleus tridentatus)
C;Species: Tachypleus tridentatus
C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 18-Jun-1993
C;Accession: A30068
R;Nakamura, T.; Furunaka, H.; Miyata, T.; Tokunaga, F.; Muta, T.; Iwanaga, S.; Niwa, M.;
J. Biol. Chem. 263, 16709-16713, 1988
A;Title: Tachyplesin, a class of antimicrobial peptide from the hemocytes of the horsesh
A;Reference number: A30068; MUID:89034158
A;Accession: A30068
A;Molecule type: protein
A;Residues: 1-17 <NAK>

Query Match 37.8%; Score 29.5; DB 2; Length 17;
Best Local Similarity 53.8%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

```
Qy       1 RVCAAGRGETCYR 13
           |||  |    |||
Db       5 RVCYRG---ICYR 14
```

RESULT 13
JU0123
tachyplesin II - horseshoe crab (Tachypleus tridentatus)
C;Species: Tachypleus tridentatus
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 11-Jul-1997
C;Accession: JU0123
R;Miyata, T.; Tokunaga, F.; Yoneya, T.; Yoshikawa, K.; Iwanaga, S.; Niwa, M.; Takao, T.;
J. Biochem. 106, 663-668, 1989
A;Title: Antimicrobial peptides, isolated from horseshoe crab hemocytes, tachyplesin II,
A;Reference number: A91914; MUID:90110066
A;Accession: JU0123
A;Molecule type: protein
A;Residues: 1-17 <MIY>
C;Comment: The peptide is one of the antimicrobial peptides found in the Japanese horses
C;Keywords: amidated carboxyl end
F;3-16,7-12/Disulfide bonds: #status predicted
F;17/Modified site: amidated carboxyl end (Arg) #status experimental

Query Match 37.8%; Score 29.5; DB 2; Length 17;
Best Local Similarity 53.8%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

```
Qy       1 RVCAAGRGETCYR 13
           |||  |    |||
Db       5 RVCYRG---ICYR 14
```

RESULT 14
S59504
ferric pseudobactins receptor protein RF2 - Pseudomonas putida (fragment)
C;Species: Pseudomonas putida
C;Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 10-Jul-1998
C;Accession: S59504
R;Koster, M.; Ovaa, W.; Bitter, W.; Weisbeek, P.
Mol. Gen. Genet. 248, 735-743, 1995
A;Title: Multiple outer membrane receptors for uptake of ferric pseudobactins in Pseu
A;Reference number: S59503
A;Accession: S59504
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-185 <KOS>
C;Superfamily: tonB-dependent receptor carboxyl-terminal homology
F;1-185/Domain: tonB-dependent receptor carboxyl-terminal homology (fragment) <TNC>

Query Match 44.9%; Score 35; DB 2; Length 185;
Best Local Similarity 41.7%; Pred. No. 41;
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 RVCAAGRGETCY 12
           : |:|| : :||
Db      30 KGCGAGPASACY 41
```

RESULT 15
JX0124
tachyplesin I precursor - horseshoe crab (Carcinoscorpius rotundicauda)
C;Species: Carcinoscorpius rotundicauda
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 11-Jul-1997
C;Accession: JX0124
R;Muta, T.; Fujimoto, T.; Nakajima, H.; Iwanaga, S.
J. Biochem. 108, 261-266, 1990
A;Title: Tachyplesins isolated from hemocytes of southeast Asian horseshoe crabs (Car
ssing intermediate of its precursor.
A;Reference number: JX0124; MUID:91035357
A;Accession: JX0124
A;Molecule type: protein
A;Residues: 1-19 <MUT>
A;Experimental source: hemocyte
C;Keywords: amidated carboxyl end
F;1-17/Product: tachyplesin I #status experimental <MAT>
F;3-16,7-12/Disulfide bonds: #status predicted
F;17/Modified site: amidated carboxyl end (Arg) (amide in mature form from following

Query Match 37.8%; Score 29.5; DB 2; Length 19;
Best Local Similarity 53.8%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

```
Qy       1 RVCAAGRGETCYR 13
           |||  |    |||
Db       5 RVCYRG---ICYR 14
```

Search completed: May 4, 1999, 08:18:17
Job time: 6969 sec